

UNIVERSIDADE DE LISBOA  
FACULDADE DE CIÊNCIAS  
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



## **Modelação da elasticidade do preço na renovação Automóvel**

Carolina Ferreira Duarte Da Silva

**Mestrado em Matemática Aplicada à Economia e Gestão**

Trabalho de Projeto orientado por:  
Professor Doutor João José Ferreira Gomes  
Joana Rita Barradas Garraio

2018



*“Education is the most powerful weapon which you can use to change the world”.*

Nelson Mandela (1918-2013)



# Agradecimentos

À minha orientadora da Ageas, Joana Garraio, por tudo o que me ensinou desde o dia em que entrei na companhia, pelo apoio e disponibilidade demonstrada em fazer parte deste projeto.

À professora Teresa Alpuim e ao professor João Gomes.

À Ageas Seguros, em especial ao Sjoerd Smeets e à Paula Santos, por me terem proporcionado a realização deste projeto, que me permitiu adquirir novos conhecimentos essenciais no mercado segurador. Obrigada por me terem proporcionado estes 2 anos na companhia, que tanto me ensinaram e continuam a ensinar.

Aos meus pais, pela oportunidade que me proporcionaram, sem eles este projeto não teria sido possível, pelo incentivo e apoio incondicional na minha formação.

Aos meus irmãos, à minha família e amigos por todo o apoio durante todos estes anos de faculdade.

Ao Vitor, pela paciência e apoio demonstrado ao longo destes anos.

Por fim, mas não menos importantes, aos meus colegas da AGEAS, em especial à Andreia Ginja e à Inês Baptista, pela amizade, ajuda e conselhos.

# Resumo

Em Portugal tem-se verificado um aumento da competitividade no mercado segurador automóvel, levando as seguradoras a considerar uma abordagem de tarifação mais assente na procura. Desta forma as seguradoras pretendem estudar os comportamentos dos clientes. Uma vez que o cliente do seguro automóvel é muito sensível às alterações de preço, é necessário analisar quais as ações que podem resultar em reações disruptivas. Assim, é crucial saber quem são os nossos clientes, qual o seu comportamento e quais as ações que podem ser tomadas sem prejudicar o seu relacionamento com a empresa.

Este projeto tem como principais objetivos modelar a taxa de anulação na renovação do seguro automóvel e analisar como diversas variáveis influenciam as renovações. Utilizou-se a regressão logística, um caso particular dos Modelos Lineares Generalizados, para modelar a variável de resposta binária (Anula/Não Anula).

Sendo a variação do prémio a única variável que a seguradora pode controlar diretamente, obter informação sobre a elasticidade do preço de cada cliente permitirá à seguradora tomar melhores decisões, com o intuito de aperfeiçoar o equilíbrio entre o grau de satisfação do cliente e a rentabilidade da companhia.

Para a construção da base utilizaram-se as renovações que tiveram lugar entre janeiro 2016 a junho 2017. No ano de 2017 apenas se considerou o 1º semestre devido à proximidade da data de renovação. A base é composta por apólices de clientes individuais, em que o objeto seguro é um veículo ligeiro, misto ou caminheta. O modelo não contempla apólices frota, protocolos nem ordens profissionais.

Por fim, de forma a avaliar a capacidade do modelo em prever que clientes irão anular as suas apólices, utilizaram-se vários critérios de otimização de ponto de corte, de modo a encontrar o valor que resulta na melhor capacidade discriminatória global.

**Palavras-chave:** Seguro automóvel, Regressão Logística, Taxa de anulação

# Abstract

It has been observed in Portugal an increase in the competitiveness of the motor insurance market, leading insurance companies to consider a pricing tariff based in demand. This way, insurance companies intend to study its clients' behaviors. The motor insurance client is highly sensitive to changes in price, so it is necessary to analyze which actions may result in disruptive reactions. That being said, it is crucial to know who our clients are, what is their behavior and which are the actions that can be taken without harming the relationship between the company and the client.

This project has as main goals to model the cancelation rate at renewal of the motor insurance and to analyze how several variables influence the renewals. Logistic regression has been used, a particular case of the Generalized Linear Models, to model the binary variable (cancel/does not cancel).

Since the variation in the premium is the only variable the insurance company can control directly, obtaining information about each client's price elasticity will allow the company to make more suitable decisions, with the aim to perfect the balance between the client's satisfaction and the company's profitability.

For the construction of the database, the renewals that took place between January 2016 and June 2017 were used. For 2017, we only considered the first semester due to the proximity of the renewal date. The database includes personal clients' policies, and the insured object is a passenger vehicle, mixed or "*caminheta*". The model does not include fleet policies, protocols and professional orders.

Besides that, in order to assess the model's capacity to predict which clients will cancel their policies, several threshold optimization criteria have been used, aiming to find the value that provides the best overall discriminatory capacity.

**Key words:** Motor insurance, Logistic regression, Cancellation rate.

# Conteúdo

Glossário e Abreviatura .....	5
Introdução .....	7
CAPÍTULO I – Atividade Seguradora .....	8
1.1. História do ramo segurador em Portugal .....	8
1.2. Grupo AGEAS .....	8
1.2.1. AGEAS .....	8
1.2.2. AGEAS Portugal .....	9
1.3. Atividade Seguradora .....	9
1.3.1. Contrato de Seguro .....	9
1.3.2. Classificação e Divisão dos Seguros .....	10
1.4. Seguro Automóvel .....	10
1.4.1. Seguro obrigatório de Responsabilidade Civil .....	11
1.4.2. Seguro facultativos – Danos próprios .....	11
1.4.3. Sistema de Bónus Malus .....	11
1.4.4. Processo de renovações automóvel .....	12
CAPÍTULO II – Enquadramento Teórico .....	13
2.1 Modelo linear generalizado .....	13
2.1.1. Família Exponencial .....	13
2.1.2. Estimação de parâmetros .....	14
2.1.3. Desvio ou análise do erro .....	14
2.1.4. Critério de Informação de Akaike e Bayesiano .....	15
2.2 Regressão Logística .....	15
2.3 Teste de Wald .....	16
2.4 Qualidade do ajustamento das variáveis binárias .....	16
2.4.1. Curva de ROC .....	17
2.5 Método de seleção de Backward (Método da seleção regressiva) .....	18
CAPÍTULO III – Construção do modelo .....	19
3.1. Preparação e listagem das variáveis em estudo .....	19
3.2. Análise Descritiva .....	20
3.2.1. Variáveis associadas ao cliente .....	21
3.2.2. Variáveis associadas à apólice .....	24



3.2.3.	Variáveis associadas ao veículo .....	33
3.2.4.	Variáveis Demográficas .....	37
3.2.5.	Variáveis associadas ao preço .....	38
3.2.6.	Variáveis associadas a outros .....	39
3.3.	Modelização .....	40
3.3.1.	Correlações.....	40
3.3.2.	Métodos de seleção de variáveis .....	41
3.3.3.	Modelo final .....	49
3.4.	Avaliação do modelo.....	50
3.4.1.	Curva de ROC .....	50
3.4.2.	Sensibilidade e Especificidade .....	51
CAPÍTULO IV – Conclusão .....		53
Referências bibliográficas .....		55

## Lista de tabelas

Tabela 2.1- Matriz de Confusão .....	16
Tabela 2. 0.2 – Classificação dos valores de AUC.....	18
Tabela 3. 1 – Variáveis em estudo.....	20
Tabela 3. 2 – Variáveis altamente correlacionadas.....	40
Tabela 3.3 – 1ª Iteração do método de Backwards .....	41
Tabela 3.4 - 2ª Iteração do método de Backwards .....	42
Tabela 3.5 - 3ª Iteração do método de Backwards .....	42
Tabela 3.6 - 4ª Iteração do método de Backwards .....	43
Tabela 3.7 - 5ª Iteração do método de Backwards .....	43
Tabela 3. 8 - 6ª Iteração do método de Backwards .....	44
Tabela 3.9 - 7ª Iteração do método de Backwards .....	44
Tabela 3.10 - 8ª Iteração do método de Backwards .....	45
Tabela 3. 11 - 1ª Iteração do teste de Wald.....	45
Tabela 3.12 - 2ª Iteração do teste de Wald.....	46
Tabela 3.13 - 2ª Iteração do teste de Wald.....	46
Tabela 3.14 - 4ª Iteração do teste de Wald.....	47
Tabela 3.15 - 5ª Iteração do teste de Wald.....	47
Tabela 3.16 - 6ª Iteração do teste de Wald.....	48
Tabela 3.17 - 7ª Iteração do teste de Wald.....	48
Tabela 3. 18 - 8ª Iteração do teste de Wald.....	49
Tabela 3.19 – Matriz confusão da base modelo .....	52
Tabela 3.20 – Matriz de confusão da base teste.....	52

## Lista de figuras

Figura 1.1 – Países onde a AGEAS se encontra presente .....	9
Figura 1.2 – Modelo de renovações na AGEAS .....	12
Figura 2.1 – Curva de ROC.....	17
Figura 3.1 – Grupos das variáveis em estudo.....	19

## Lista de gráficos

Gráfico 3. 1 – Taxa de anulação por Idade do tomador.....	21
Gráfico 3.2 - Taxa de anulação por Idade do Condutor .....	21
Gráfico 3. 3 – Taxa de anulação por Idade da Carta .....	22
Gráfico 3. 4 - Taxa de anulação por Tipo de Cliente Automóvel.....	22
Gráfico 3. 5 - Taxa de anulação por Tipo Cliente Não-Vida.....	23
Gráfico 3. 6 - Taxa de anulação por Tipo Cliente Vida .....	23
Gráfico 3. 7 – Tipo Cliente produto .....	24
Gráfico 3.8 - Taxa de anulação por Ano de Subscrição .....	24
Gráfico 3. 9 - Taxa de anulação por Antiguidade da apólice.....	25
Gráfico 3. 10 - Taxa de anulação por Mês de renovação .....	25
Gráfico 3.11 - Taxa de anulação por Modalidade .....	26
Gráfico 3.12 - Taxa de anulação por Antiguidade do Agente .....	26
Gráfico 3.13 - Taxa de anulação por Canal do Agente .....	27
Gráfico 3.14 - Taxa de anulação por Forma de Cobrança .....	27
Gráfico 3.15 - Taxa de anulação por Forma de Pagamento .....	28
Gráfico 3.16 - Taxa de anulação por Capital da Apólice (1) .....	28
Gráfico 3.17 - Taxa de anulação por Capital da Apólice (2) .....	29
Gráfico 3. 18 - Taxa de anulação por Tipo de Pack .....	29
Gráfico 3.19 - Taxa de anulação por Pack da Apólice .....	30
Gráfico 3.20 - Taxa de anulação por Tarifa inicial da Apólice .....	30
Gráfico 3. 21 – Taxa de anulação por sinistros agravantes nas últimas 5 anuidades .....	31
Gráfico 3. 22 – Taxa de anulação por sinistros agravantes nas últimas 3 anuidades .....	31
Gráfico 3.23 - Taxa de anulação por sinistros agravantes na última anuidade .....	31
Gráfico 3. 24 - Taxa de anulação por Alteração do prémio .....	32
Gráfico 3. 25 - Taxa de anulação por Alteração do escalão de bónus .....	32
Gráfico 3. 26 - Taxa de anulação por Categoria do veículo.....	33
Gráfico 3. 27 - Taxa de anulação por Marca do veículo .....	33
Gráfico 3. 28 - Taxa de anulação por Idade do veículo .....	34
Gráfico 3.29 - Taxa de anulação por número de CC.....	34
Gráfico 3. 30 - Taxa de anulação por número de CV.....	35
Gráfico 3.31 - Taxa de anulação por Tara do veículo .....	35
Gráfico 3.32 - Taxa de anulação por Tipo de Combustível.....	36
Gráfico 3.33 - Taxa de anulação por Peso Bruto do veículo .....	36
Gráfico 3.34 - Taxa de anulação por Valor em novo do veículo .....	37

Gráfico 3. 35 - Taxa de anulação por Urban Rural .....	37
Gráfico 3. 36 - Taxa de anulação por Distrito (1) .....	38
Gráfico 3.37 - Taxa de anulação por Distrito (2) .....	38
Gráfico 3.38 - Taxa de anulação por Intervalo de Variação Relativa .....	39
Gráfico 3. 39 - Taxa de anulação por Intervalo de Variação Relativa na anuidade anterior .....	39
Gráfico 3. 40 - Taxa de anulação por Cartão Ageas .....	40
Gráfico 3.41 – Curva de ROC da base modelo .....	50
Gráfico 3.42 – Curva de ROC da base teste.....	51
Gráfico 3. 43 – Sensibilidade e Especificidade da base modelo.....	51
Gráfico 3.44 - Sensibilidade e Especificidade da base teste .....	51

# Glossário e Abreviatura

**AGE** - Agentes exclusivos, que apenas trabalham com a Ageas.

**Agente** – O Agente é um mediador que pode exercer atividade para uma ou mais companhias de seguros, onde angaria clientes, apresenta soluções e propostas, que posteriormente são traduzidas em contratos de seguro.

**APS** – Associação Portuguesa de Seguradores, defende e promove os interesses das seguradoras.

**Apólice de seguro** – Contrato de seguro efetuado entre a entidade seguradora e o tomador da apólice

**ASF** - Autoridade de Supervisão de Seguros e Fundos de Pensões, entidade responsável pela regulação e supervisão das seguradoras, resseguradoras, fundos de pensões e mediação de seguros.

**AVA** - Assistência em viagem automóvel

**Beneficiário** – Pessoa singular ou coletiva beneficiária do capital proveniente do contrato ou de uma operação de capitalização (indenização ou entrega de capital).

**BM** - O sistema Bónus/Malus consiste numa escala progressiva de agravamentos por sinistro e de bónus na ausência do mesmo.

**Bónus** – O tomador terá direito a uma redução do prémio puro, quando durante 2 anuidades consecutivas não houver qualquer participação de sinistro. Só são considerados sinistros, aqueles que tiverem lugar ao pagamento de indemnizações.

**B2B** - Banca e Parcerias

**CAMEO** – Base de dados externa à companhia, com a informação por código postal.

**CCC** - Choque, Colisão ou Capotamento

**Dano Corporal** – Toda a lesão ou sequela resultante no corpo de uma pessoa vítima de um trauma corporal (físico, psíquico ou social).

**Dano Material** – Toda a lesão causada aos interesses de outrem e que venham a causar diminuição patrimonial a esse outrem.

**DP** - Danos Próprios - O seguro de Danos Próprios abrange prejuízos sofridos pelo veículo seguro independentemente de quem é o responsável pelo acidente, mesmo que seja o próprio condutor.

**Franquia** – Importância a cargo do cliente em caso de sinistro, definida no contrato do seguro.

**FR** - Furto ou Roubo

**GLM** - Generalized linear model (Modelos Lineares Generalizados-MLG)

**IRE** - Incêndio, Raio ou Explosão

**Malus** – O tomador terá um aumento do prémio na renovação do contrato por ocorrência de sinistros.

**MM** - Agente multimarca, são agentes que trabalham com diversas seguradoras.

**Perda total** – É considerado perda total quando ocorre destruição total do veículo ou objeto seguro, cujos custos de reparação são superiores ao valor declarado antes do sinistro.

**Prémio comercial<sup>1</sup>** – Custo técnico das coberturas do contrato, este prémio não considera os custos relacionados com a emissão do contrato, de aquisição, de gestão ou de cobrança.

**Prémio total** – O prémio é a contrapartida da cobertura acordada e inclui tudo o que seja contratualmente devido pelo tomador do seguro, nomeadamente os custos da cobertura do risco, os custos de aquisição, de gestão e de cobrança, custos de fracionamento, custo de Apólice/ata adicional, carta verde e dos encargos fiscais e parafiscais a suportar pelo Tomador do seguro.

**Private** - Agentes exclusivos, colaboradores AGEAS

**Retenção** – Aceitação da renovação do contrato de seguro por parte do segurado.

**Risco** – Corresponde à possibilidade de ocorrência de determinado acontecimento indesejável, aleatório, fortuito, com consequências negativas e alheio à vontade dos intervenientes. Por outro lado, corresponde ao conjunto de perigos cobertos pela apólice de seguro;

**Segurado** – É a entidade ou pessoa que fica segurada no contrato. No caso do seguro automóvel, o segurado é o condutor do veículo.

**Seguradora** – Entidade legalmente constituída para assumir e gerir riscos, devidamente especificados no contrato de seguro.

**Segurnet** – Plataforma nacional com o histórico das matrículas e respetiva sinistralidade.

**Sinistro** – Evento ou série de eventos, resultantes de uma mesma causa, suscetível de fazer funcionar as garantias de um ou mais contratos de seguro.

**Terceiro** – Vítima resultante de um sinistro provocado pelo segurado, cujos danos que sofreu são suscetíveis de serem cobertos pela seguradora.

**Tomador do seguro** – É a pessoa que celebra o contrato de seguro com a seguradora, é aquele que se compromete a cumprir a obrigação contratual.

**RC** - Responsabilidade Civil

**Valor em Novo** – Valor do veículo quando este é novo.

**Valor seguro** – No caso das apólices que têm cobertura de danos próprios, o valor seguro do veículo é o valor utilizado para calcular a indemnização em caso de perda total.

---

<sup>1</sup> Conforme a página online da ASF

# Introdução

Este relatório é referente ao estágio realizado na Ageas Seguros na equipa de *Pricing & Business Analytics*, inserida na direção Técnica e Oferta Não Vida liderada pelo Dr. Sjoerd Smeets. Durante este estágio abordou-se exclusivamente o ramo Automóvel.

A direção Técnica e Oferta é uma das áreas mais importantes numa empresa seguradora. É esta direção que cria, desenvolve e implementa todas as estratégias para ajustar a oferta e o preço dos produtos que a seguradora oferece aos clientes. Por sua vez, a equipa de *Pricing & Business Analytics* tem como intuito analisar o preço e gerir o portfólio.

Quando as seguradoras Não-Vida desenvolvem as suas estratégias financeiras, têm em consideração dois momentos cruciais do seu relacionamento com os clientes. O primeiro é o momento de aquisição dos riscos e assinatura do contrato (conversão da proposta). O segundo, que é onde se foca este projeto, é o momento de renovação desse contrato.

Em Portugal, como na maioria dos países, o seguro de responsabilidade civil é obrigatório, o que levou a um mercado de seguros automóvel muito competitivo, com o prémio médio a diminuir sistematicamente ao longo de 10 anos até 2014 (Associação Portuguesa de Seguradores, 2015). O aumento do nível de concorrência obrigou as seguradoras a perceberem quem são os seus clientes, como se comportam e quais as ações que as seguradoras podem e devem tomar durante o período de renovação de forma a evitar perder os bons clientes, mantendo ou aumentando o lucro.

Com base neste cenário surgiu este projeto, tendo como principais objetivos modelar a taxa de anulação do seguro automóvel na companhia assim como analisar quais as variáveis que influenciam a renovação.

Uma vez que a alteração de prémio é a única variável que a seguradora tem controlo direto, este modelo pode ajudar a empresa a compreender como a elasticidade do preço varia de cliente para cliente e qual a variação de prémio mais adequada de forma a reter os seus clientes. A obtenção deste conhecimento permitirá à seguradora tomar melhores decisões a fim de alcançar o equilíbrio de satisfação e a rentabilidade do cliente.

Neste projeto iremos utilizar os Modelos Lineares Generalizados, em particular o caso de regressão logística, para modelar a variável resposta.

Para a construção da base de dados utilizou-se o software SAS Enterprise Guide, no entanto para construir e testar o modelo de regressão logística utilizou-se o Emblem.

Relativamente à estrutura deste documento, no Capítulo 1 é realizada uma abordagem à atividade seguradora, à seguradora onde este projeto foi realizado e por fim aborda-se o ramo automóvel. No capítulo 2, é realizado um enquadramento teórico, começando por uma visão geral dos Modelos Lineares Generalizados, seguido da regressão logística e uma descrição de várias medidas de desempenho da variável binária. No capítulo 3, falamos sobre a construção do modelo, apresentamos as variáveis utilizadas para modelar e analisamos diversos critérios de seleção de variáveis. Por fim, analisamos o modelo final utilizado, a curva de ROC e a matriz de confusão.

# CAPÍTULO I – Atividade Seguradora

## 1.1. *História do ramo segurador em Portugal*<sup>2</sup>

Portugal é um dos impérios mais antigos da Europa, sendo um dos principais exploradores do caminho marítimo. Em 1293 o Rei D. Dinis criou a primeira forma de Seguro em Portugal. Este seguro começou por ser dedicado exclusivamente aos riscos associados às descobertas marítimas, por isso D. Dinis ordenou que fosse celebrado um acordo entre mercadores, tendo como objetivo o pagamento de certas quantias (Prémio) sobre as embarcações. Estes pagamentos eram calculados de acordo com o porte da embarcação e do comércio. Estes montantes serviam para fazer face aos acidentes (Sinistros), por perdas de navios e mercadorias. Em 1370, no reinado de D. Fernando, foi promulgada a primeira lei sobre seguros, dando continuidade ao trabalho desenvolvido por D. Dinis. Dez anos mais tarde foi criada a Companhia das Naus, que funcionava como companhia de seguros. Todos os navios tinham que ser registados, pagando uma percentagem dos lucros de cada viagem. Estes fundos serviam para cobrir os prejuízos das embarcações que se afundassem ou sofressem avarias. A criação da Companhia das Naus constituiu uma medida muito importante, que se revelaria precursora da expansão ultramarina portuguesa. Percebendo a importância que os seguros tinham, no ano de 1383 surge em Portugal a primeira Lei Nacional sobre Seguros.

A partir do século XV começam a surgir os primeiros Seguros de Vida.<sup>3</sup> No ano da Batalha de Alcácer Quibir (1578), foi criado em Portugal o cargo de Corretor de Seguros, com o objetivo de regulamentar e fiscalizar todas as operações, acabando com os abusos que tinham originado prémios exageradamente altos, e dificuldades extremas nos casos de regularização dos sinistros.

No ano 1791 foi criada a primeira companhia de seguros portuguesa, designada por “Companhia Permanente de Seguros”, tendo um capital inicial de 60 contos de réis. Esta primeira companhia impulsionou a criação de inúmeras companhias nos anos seguintes. Desta forma, a área seguradora foi sempre evoluindo ao longo dos séculos.

## 1.2. *Grupo AGEAS*

### 1.2.1. *AGEAS*<sup>4</sup>

A AGEAS é uma companhia seguradora internacional, sediada em Bruxelas, com 190 anos de experiência e de conhecimento. A companhia começou por se chamar Fortis, tornando-se Ageas em 2008.

Está presente em 13 países da Europa e da Ásia - Bélgica, Reino Unido, França, Itália, Portugal, Turquia, China, Malásia, Índia, Tailândia, Vietname e Filipinas, sendo um dos maiores grupos seguradores europeus e líder na Bélgica. Encontra-se entre os principais *players* na maioria dos países em que está presente.

---

<sup>2</sup> Conforme a página online [historiadoseguro.com](http://historiadoseguro.com)

<sup>3</sup> De acordo com as notas teóricas Figueiredo, P, Atividade Seguradora

<sup>4</sup> De acordo com a página online da empresa



Figura 1.1 – Países onde a AGEAS se encontra presente

A empresa propõe soluções Vida e Não Vida a milhões de clientes individuais e empresariais. A AGEAS ajuda os seus Clientes a gerir, antecipar e proteger-se contra os riscos através de um conjunto alargado de produtos desenhados para responder às suas necessidades, presentes e futuras.

Reconhecida pela sua forte experiência em matéria de parcerias, a Ageas desenvolveu acordos de longa duração com instituições financeiras e distribuidores locais de referência pelo mundo inteiro, de forma a garantir proximidade com os seus Clientes.

A imagem de marca da empresa tem um significado próprio e específico, os círculos desenhados à mão são um símbolo da energia e impaciência para concretizar e realizar. As cores sugerem uma empresa multicultural capaz de se adaptar a ambientes em mudança.

### ***1.2.2. AGEAS Portugal***

A Ageas surge em Portugal no início do século XXI após variadas fusões ao longo dos anos. O grupo entrou no mercado segurador Português em 2005, através de uma *joint venture* com o Banco Millennium BCP, sendo uma empresa líder no *ranking* segurador português e operando com reconhecidas marcas como a Ocidental e a Médis.

Em abril de 2016 o grupo AGEAS adquiriu a AXA Portugal, denominando-se Ageas Seguros e reforçando a sua aposta de investimento em Portugal, completando os seus canais de distribuição já existentes com uma rede de Mediadores Profissionais e um canal de venda direta. A Ageas Seguros opera nos segmentos Vida e Não Vida.

## ***1.3. Atividade Seguradora***

### ***1.3.1. Contrato de Seguro***

O seguro é a operação através da qual o tomador do seguro, entidade (individual ou coletiva) que contrata o seguro com a seguradora, transfere para a seguradora determinados riscos aleatórios. Estes riscos correspondem à possibilidade de ocorrência de determinados acontecimentos indesejáveis, aleatórios, fortuitos, com consequências negativas e alheias à vontade dos intervenientes. Por outro lado, corresponde ao conjunto de perigos cobertos pela apólice do seguro.

Em Portugal, juridicamente, a sistematização do contrato de seguro só ocorre em 1888 por via da publicação do primeiro Código Comercial.



Deste modo, o contrato de seguro é definido como um acordo através do qual o segurador assume a cobertura de determinados riscos, comprometendo-se a satisfazer as indemnizações ou a pagar o capital seguro em caso de sinistro (realização total ou parcial do risco), nos termos acordados. Em contrapartida, a pessoa ou entidade que celebra o seguro (tomador do seguro) fica obrigada a pagar ao segurador o prémio correspondente.

Alguns contratos necessitam da intervenção de várias seguradoras, devido à sua dimensão, natureza ou característica, o que proporciona um nível de risco demasiado elevado. Para este tipo de riscos criou-se os contratos de co-seguro. Este é um contrato celebrado entre várias seguradoras, onde uma é líder, sendo esta a seguradora com maior responsabilidade no contrato, e nela fica estabelecida a participação de cada co-seguradora no total da quantia segurada.

### ***1.3.2. Classificação e Divisão dos Seguros***

Os seguros estão divididos em dois grandes grupos - seguros do ramo vida e os seguros do ramo não vida.

Os seguros de ramo Vida são os seguros de vida risco, geralmente cobrem situações de morte e invalidez, e os seguros financeiros (PPR-Planos de Poupança Reforma).

No ramo Não Vida existem vários seguros, sendo os mais usuais e conhecidos os seguintes:

- Automóvel;
- Acidentes de trabalho;
- Acidentes pessoais;
- Saúde;
- Patrimoniais (Multirriscos habitação e multirriscos industrial e comércio);
- Entre muitos outros.

Existe um conjunto de seguros obrigatórios por lei que cobrem uma extensa lista de atividades. Os mais conhecidos e os que estão mais diretamente ligados ao nosso dia-a-dia são:

- Acidentes de trabalho;
- Automóvel – Responsabilidade civil;
- Caçador – Responsabilidade civil;
- Escolar - Acidentes pessoais;
- Entre outros.

### ***1.4. Seguro Automóvel***

O seguro automóvel é o seguro mais popular e o mais representado na estrutura das carteiras de qualquer seguradora portuguesa no que respeita a seguros do ramo não vida. É também um dos seguros com maior taxa de sinistralidade. Ainda assim, cada vez mais existe uma maior concorrência entre as seguradoras, oferecendo soluções cada vez mais baratas e que acabam por diminuir as margens de lucro deste seguro para as companhias.

Este é um produto chave para as companhias, uma vez que a empresa onde um cliente subscrever um seguro automóvel será, em princípio, a empresa onde o cliente subscreverá os seus

outros seguros. Podemos mesmo considerar o seguro automóvel como a porta de entrada para uma relação mais duradoura com o cliente.

#### ***1.4.1. Seguro obrigatório de Responsabilidade Civil***

Qualquer veículo a motor que se desloque na via pública, tem que se encontrar obrigatoriamente ao abrigo de um seguro automóvel através da sua cobertura de responsabilidade civil, ou seja, uma cobertura que permite que em caso de danos provocados a terceiros, sejam eles corporais ou materiais, estes prejuízos possam ser devidamente indemnizados pela seguradora aos lesados. A cobertura de RC obrigatória é imposta pelo governo, sendo que a partir de junho de 2017 o capital de responsabilidade civil passou de 6.000.000€ para 7.290.000€ (6.070.000€ para RC danos corporais e 1.220.000€ RC danos materiais).

O capital de responsabilidade civil obrigatório para transporte coletivo de passageiros também foi alterado, passando de 12.000.000€ para 14.580.000€ (12.140.000€ para RC danos corporais e 2.440.000€ RC danos materiais).

No entanto, existe algumas exclusões nesta cobertura, como por exemplo:

- Condutor e tomador do seguro;
- Danos causados no veículo próprio, bens transportados no veículo seguro;
- Danos causados a terceiros em consequência de operações de carga e descargas.

#### ***1.4.2. Seguro facultativos – Danos próprios***

Para além do seguro de responsabilidade civil, existe um conjunto de coberturas pelo qual o tomador do seguro poderá optar e que tornará o seu seguro mais completo e colocará o seu veículo a salvo de um conjunto de riscos adicionais.

São diversas as coberturas de danos próprios existentes. Algumas delas já são comercializadas no pack de responsabilidade civil, sendo elas a quebra isolada de vidros (QIV) e a assistência em viagem automóvel (AVA). As coberturas de Choque, Colisão e Capotamento (CCC), Furto ou Roubo (FR) e Incêndio, Raio e Explosão (IRE) são as coberturas mais comercializadas quando se adquire um pack de danos próprios. Contudo, existem muitas outras coberturas associadas a fenómenos da natureza, riscos sociais e veículos de substituição.

Nas coberturas facultativas temos ainda a cobertura de RC facultativa. Esta cobertura é um complemento à cobertura de RC obrigatória, e tem um capital seguro de 50.000.000€.

#### ***1.4.3. Sistema de Bónus Malus***

O sistema de Bónus Malus (BM) apenas é aplicado no seguro automóvel, no entanto nem todas as coberturas são afetadas, sendo as seguintes as afetadas: Responsabilidade Civil (obrigatória e facultativa), Choque, Colisão ou Capotamento e Incêndio, Raio ou Explosão, Furto ou Roubo, Só Colisão e Solução Ano Seguro.

Este sistema é uma forma das companhias beneficiarem ou agravarem os prémios dos clientes consoante o histórico de sinistros que estes tenham. Quando o cliente tem um bónus é aplicado um desconto percentual nas coberturas afetadas pelo sistema, caso o cliente tenha um malus é aplicado um aumento percentual.

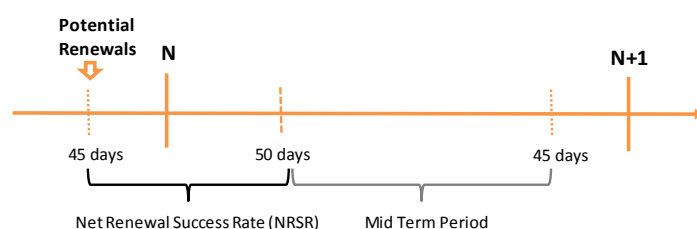
A percentagem do sistema de BM difere de companhia para companhia, e a atribuição de Bónus/Agravamento depende do número de anos de vigência do contrato e do número de sinistros registados (histórico comprovado através da Segurnet, caso seja uma transferência externa, ou do histórico da Ageas, caso seja uma transferência interna).

#### ***1.4.4. Processo de renovações automóvel***

O processo de renovações é dos processos mais importantes numa companhia de seguros. É neste processo que a empresa determina os preços para a anuidade seguinte do contrato. Para este processo são tidos em conta vários fatores, e a forma de processar as renovações varia de companhia para companhia.

O processo de renovações, na Ageas, considera todas as apólices que estejam em vigor 45 dias antes da renovação, ou seja, 45 dias antes da potencial renovação do contrato. Nessa altura a empresa envia ao cliente uma comunicação com o valor do novo prémio para a anuidade seguinte. A situação da apólice é verificada 50 dias após a data de renovação, se a apólice estiver em vigor considera-se que a apólice foi renovada com sucesso. Se a apólice estiver anulada considera-se que a apólice não foi renovada com sucesso. Nos casos em que a forma de pagamento é mensal, caso o cliente pague a primeira mensalidade e não pague a segunda, consideramos que houve um cancelamento do contrato.

No entanto a anulação do contrato pode ocorrer após os 50 dias. Nesse caso, e como essas anulações não são causadas pela variação do prémio, não são consideradas na nossa análise como anulações por causa do processo de renovação. Para terminar, define-se:



*Figura 1.2 – Modelo de renovações na AGEAS*

***Potential Renewals*** – Apólices em vigor no momento do processo de renovação, sendo enviada uma carta ao cliente com a informação da renovação.

***Net Renewal Success Rate (NRSR)*** – Percentagem de *Potential Renewals* que realmente renovaram no período de 50 dias (após o início da nova anuidade). Uma apólice é considerada renovada se e só se cumpre as seguintes condições:

- A primeira prestação é paga;
- Não existe data de cancelamento após 50 dias do início da nova anuidade.

***Mid Term Cancellation Rate*** – É a percentagem de apólices canceladas durante o *Mid Term Period* (50 dias depois da renovação até 45 dias antes da renovação seguinte).

# CAPÍTULO II – Enquadramento Teórico

Neste capítulo pretende-se dar a conhecer os modelos lineares generalizados, também conhecidos por Generalized Linear Models (GLM). Falaremos em especial da regressão logística, apropriada para modelar variáveis binárias, incluindo algumas considerações aos testes do modelo.

## 2.1 Modelo linear generalizado

Apresentados por Nelder e Wedderburn (1972), os modelos lineares generalizados (GLM) correspondem a uma síntese de vários modelos desenvolvidos anteriormente, vindo assim unificar, tanto no ponto de vista teórico como conceptual, a teoria da modelação estatística desenvolvida até 1972 (Turkman&Silva (2000)).

O estudo destes modelos e da sua aplicabilidade tem-se vindo a desenvolver ao longo do tempo e estes modelos têm sido cada vez mais utilizados nos estudos atuariais. Os GLM são utilizados para analisar os efeitos que diferentes covariáveis têm na variável resposta em estudo.

O Modelo Linear Generalizado terá que assentar sobre três componentes:

- A distribuição da variável de resposta  $Y$ , pertence à família exponencial (componente aleatória);
- $\eta = \sum_{i=1}^p x_i \beta_i$  é a combinação linear de  $p$  variáveis explicativas, onde  $x_i$  são as covariáveis e  $\beta_i$  são os parâmetros de dimensão  $p$  (componente sistemática/determinística);
- A função de ligação  $g(\mu) = \eta$ , onde  $g(\cdot)$  é uma função monótona e diferenciável em todo o seu domínio e  $\mu = E[Y]$ .

As covariáveis são incorporadas no modelo através do preditor linear, enquanto que a função ligação liga o *linear predictor* com a média da resposta. A função ligação deve ser escolhida de modo que os valores ajustados estejam dentro do domínio de  $\mu$ .

### 2.1.1. Família Exponencial<sup>5</sup>

Diz-se que uma variável aleatória  $Y$  tem distribuição pertencente à família exponencial se a sua função densidade de probabilidade (f.d.p.) ou função massa de probabilidade (f.m.p.) se puder escrever na forma:

$$f(y|\theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \quad (2.1)$$

- $\theta$ : é a forma canónica do parâmetro escalar de localização;
- $\phi$ : é um parâmetro escalar de dispersão desconhecido;

---

<sup>5</sup> De acordo com Turkman, M. Antónia e Silva, Giovani, Modelos Lineares Generalizados.

- $a(\cdot), b(\cdot)$  e  $c(\cdot, \cdot)$  são funções reais conhecidas, sendo que a distribuição de  $b(\cdot)$  não depende dos parâmetros e  $a(\phi) = \frac{\phi}{\omega}$ , onde  $\omega$  é uma constante conhecida.

Para se obter o valor médio e a variância da distribuição exponencial, deriva-se a função densidade de probabilidade,  $\ell(\theta, \phi, y) = \ln(f(y|\theta, \phi))$ , designada por função Score:

$$S(\theta) = \frac{\partial \ell(\theta; y, \phi)}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)} \quad (2.2)$$

Sabe-se que para famílias regulares se tem:

$$E(S(\theta)) = 0 \text{ e } E[S^2(\theta)] = E\left[\left(\frac{\partial \ell(\theta; y, \phi)}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \ell(\theta; y, \phi)}{\partial \theta^2}\right] \quad (2.3)$$

$$\text{Assim } S(\theta) = \frac{y - b'(\theta)}{a(\phi)} \text{ e } \frac{\partial S(\theta)}{\partial \theta} = -\frac{b''(\theta)}{a(\phi)}, \text{ onde } b'(\theta) = \frac{\partial b(\theta)}{\partial \theta} \text{ e } b''(\theta) = \frac{\partial^2 b(\theta)}{\partial \theta^2}. \quad (2.4)$$

$$\text{Desta forma obtemos, } E(Y) = \mu = b'(\theta) \text{ e } Var(Y) = a(\phi)b''(\theta). \quad (2.5)$$

A família da exponencial inclui várias distribuições, entre elas a distribuição Normal, Poisson, Gamma e Binomial, que podem ser utilizadas num modelo linear generalizado.

### 2.1.2. Estimação de parâmetros<sup>6</sup>

O método de máxima verosimilhança é utilizado para estimar os parâmetros dos GLM. Considerando a distribuição da família exponencial, a função log-verosimilhança de uma amostra aleatória  $(y_1, \dots, y_n)$  é dada por:

$$\ell(\theta, \phi; y_1, \dots, y_n) = \sum_{i=1}^n \ln f(y_i; \theta, \phi) = \sum_{i=1}^n \left[ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right]. \quad (2.6)$$

Admitindo que se verificam certas condições de regularidade os estimadores de máxima verosimilhança para  $\beta$  são obtidos como solução do sistema de equações de verosimilhança:

$$\frac{\partial \ell(\beta)}{\partial \beta_j} = 0 \Leftrightarrow \sum_{i=1}^n \frac{\partial \ell(\beta_j, y_i)}{\partial \beta_j} = 0 \Leftrightarrow \sum_{i=1}^n \left[ \frac{y_i - b'(\theta_i)}{a(\phi)} \right] \frac{\partial \theta_i}{\partial \beta_j} = 0, j = 1, \dots, p. \quad (2.7)$$

### 2.1.3. Desvio ou análise do erro

Considere-se o chamado modelo saturado, onde o número de parâmetros é igual ao número de observações. Esse modelo terá um comportamento insatisfatório quando aplicado a novos dados, no entanto pode ser comparado ao modelo atual para avaliar até que ponto o segundo modelo é perfeito. O desvio do modelo atual pode ser calculado através da seguinte fórmula:

$$D = 2\phi(\tilde{l} - \hat{l}), \quad (2.8)$$

Onde  $\tilde{l}$  é log-verosimilhança do modelo saturado e  $\hat{l}$  log-verosimilhança é log-verosimilhança do modelo atual.

---

<sup>6</sup> De acordo com Turkman, M. Antônia e Silva, Giovani, Modelos Lineares Generalizados.

Em amostras de grandes dimensões a distribuição do desvio aproxima-se à distribuição  $\chi^2(n-p)$  (sob certas condições), sendo este resultado usado habitualmente para avaliar a qualidade do modelo.

#### 2.1.4. Critério de Informação de Akaike e Bayesiano

O critério de informação de Akaike (AIC) (Akaike, 1974) é um critério de comparação de modelos, baseado na função de máxima verosimilhança, definido por:

$$AIC = -2\ell + 2p.$$

Ao comparar um conjunto de modelos diferentes, todos baseados nas mesmas observações, aquele com o AIC mais baixo é selecionado.

O Critério de Informação Bayesiano (BIC), proposto por Schwarz (1978) é dado por:

$$BIC = -2\ell + 2p * \log(n),$$

onde  $p$  é o número de parâmetros a serem estimados e  $n$  é o número de observações da amostra.

## 2.2 Regressão Logística

Considere-se uma variável aleatória binária  $Y$ , que pode tomar valores 0 e 1, sendo a probabilidade de sucesso definida por  $\mu = P[Y = 1]$ . Então  $Y \sim \text{Bernoulli}(\mu)$  e a sua função massa de probabilidade é dada por

$$f_Y(y; \theta, \phi) = \mu^y (1 - \mu)^{1-y}, \quad y=0 \text{ ou } 1. \quad (2.9)$$

Pode-se facilmente mostrar que a distribuição de Bernoulli pertence à família exponencial, escrevendo (2.9) como (2.1):

$$f_Y(y; \theta, \phi) = \mu^y (1 - \mu)^{1-y} = (1 - \mu) \left[ \frac{\mu}{1 - \mu} \right]^y = \exp \left\{ y \ln \left( \frac{\mu}{1 - \mu} \right) + \ln(1 - \mu) \right\}.$$

Assim, o parâmetro natural o  $\text{logit}(\theta) = \ln \left( \frac{\mu}{1 - \mu} \right)$ , resulta em  $\mu = \frac{e^\theta}{1 + e^\theta}$ . Sabemos também que  $\phi=1$ ,  $a(\phi) = \phi$ ,  $b(\theta) = -\ln(1 - \mu) = \ln(1 + e^\theta)$  e  $c(y, \phi) = 0$ .

Consequentemente, de (2.2) e (2.3) temos  $E[Y] = \frac{e^\theta}{1 + e^\theta} = \mu$  e  $\text{Var}[Y] = \frac{e^\theta}{1 + e^{\theta^2}} = \mu(1 - \mu)$ , respetivamente.

Uma resposta binária pode então ser modelada através de um GLM, considerando a distribuição de Bernoulli, com probabilidade de sucesso igual a  $\mu$ . Através da função logit temos a seguinte equação  $g(\mu) = \theta = \ln \left( \frac{\mu}{1 - \mu} \right)$ .

Neste projeto iremos usar um GLM com distribuição de Bernoulli e a função logit na regressão logística.

### 2.3 Teste de Wald

A estatística do teste de Wald é obtido por comparação entre a estimativa de máxima verossimilhança do parâmetro ( $\hat{\beta}_j$ ) e a estimativa do seu erro padrão. O teste de Wald univariado, realiza-se para as seguintes hipóteses:

$$H_0: \beta_j = 0 \text{ vs } H_1: \beta_j \neq 0, j = 0, \dots, p, \text{ sendo } p \text{ as variáveis a testar.}$$

A estatística do teste *Wald* para a regressão logística é dada por:  $W_j = \frac{(\hat{\beta}_j)^2}{\text{var}(\hat{\beta}_j)}$ , a hipótese nula tem distribuição Qui-quadrado com um grau de liberdade.

### 2.4 Qualidade do ajustamento das variáveis binárias

São muitos os métodos para avaliar a qualidade de ajustamento de um modelo de regressão logística. Entre alguns dos métodos estão o método de máxima verossimilhança, a matriz de confusão e a curva de ROC. Neste capítulo iremos estudar a matriz de confusão e a curva de ROC.

A classificação de um modelo irá, no caso das variáveis binárias, classificar cada observação como positiva ou negativa (1 ou 0), levando a algumas definições muito simples. Aos valores positivos que são previstos chamam-se “verdadeiros positivos” (VP) e aos que são previstos erradamente como negativos são denominados como “falsos negativos” (FN). Analogamente, os verdadeiros valores negativos podem ser previstos corretamente como sendo negativos, isto é, os “verdadeiros negativos” (VN), ou podem ser previstos como positivos e, nesse caso denominam-se “falsos positivos” (FP).<sup>7</sup>

Denotamos o número total de observações positivas por P e o número total de observações negativas por N. Os resultados observados e previstos podem ser resumidos na forma de uma matriz, a matriz de confusão, como observado na tabela abaixo:

		Valores Observados		Total
		Positivos	Negativos	
Valores Previstos	Positivos	VP	FP	VP+FP
	Negativos	FN	VN	FN+VN
Total		P=VP+FN	N=FP+VN	P+N=N.º Obs

Tabela 2.1- Matriz de Confusão

Observamos que a soma da diagonal principal nos dá o número de casos em que o modelo estava correto, A fração de casos corretos previstos é conhecida como precisão:

$$\text{Precisão} = \frac{VP+VN}{P+N}.$$

Pretende-se que os números de falsos positivos (FP) e de falsos negativos (FN) sejam os menores possíveis para que o modelo se torne o mais ajustado possível. Estes valores podem ser alterados modificando o ponto de corte, P. O ponto de corte corresponde ao ponto em que a sensibilidade se intersecta com a especificidade, é este o valor para o qual se toma a decisão de qualidade do modelo. Contudo, se, por exemplo, aumentarmos o ponto de corte, o número de FP diminui, mas aumenta o número de FN. Inversamente, se diminuirmos o valor de P do ponto de corte, diminui também o número de FN, mas aumenta o número de FP. Ou seja, por modificação

<sup>7</sup> De acordo com Alpuim, T, Modelos Lineares- Notas de apoio à disciplina, 2015.

do ponto de corte, não é possível diminuir o número de falsos positivos sem aumentar os falsos negativos, bem como o inverso.<sup>8</sup>

Podemos assim definir as seguintes probabilidades, com o intuito de apreciar a qualidade do modelo:

- ❖ **Sensibilidade (S)** – É a probabilidade de uma observação ser classificada como positiva dado que é efetivamente positiva, expressa pelo seguinte quociente;

$$S = \frac{VP}{VP + FN}$$

- ❖ **Especificidade (E)** – É a probabilidade de uma observação ser classificada como negativa dado que é efetivamente negativa, expressa pelo seguinte quociente;

$$E = \frac{VN}{VN + FP}$$

De forma a termos um modelo considerado “bom” devemos procurar manter a proporção de falsos positivos (1-E) e de falsos negativos (1-S) baixos.

#### 2.4.1. Curva de ROC

Os gráficos de *Receiver Operating Characteristics*, também conhecidos por curva de ROC, demonstram graficamente a sensibilidade versus o complementar da especificidade (1-E) por um conjunto de pontos de corte possíveis. No eixo vertical podemos ver a sensibilidade enquanto que no eixo horizontal vemos o complementar da especificidade (1-E). Consequentemente, o ponto (0,1) representa uma discriminação perfeita, enquanto os pontos (0,0) e (1,1) representam o desempenho associado aos limiares de 1 e 0, respetivamente.

A figura abaixo exemplifica uma curva ROC:

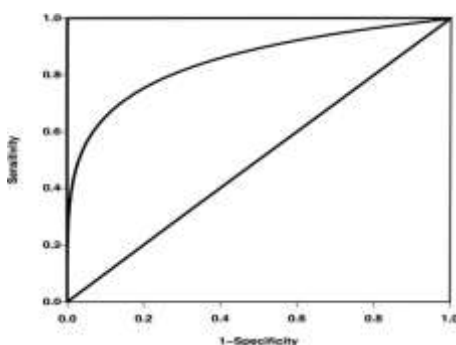


Figura 2.1 – Curva de ROC

As curvas ROC podem ser usadas para comparar modelos diferentes para os mesmos dados. Se a curva de um modelo estiver sempre acima da outra, então o primeiro modelo tem maior sensibilidade e especificidade.

Alternativamente, a área abaixo da curva de ROC (AUC) pode ser calculada, e quanto mais alto o valor melhor o modelo explica os dados. Os valores da AUC estão compreendidos

---

<sup>8</sup> De acordo com Alpuim, T, Modelos Lineares- Notas de apoio à disciplina, 2015.



entre 0 e 1 e podem ser usadas para avaliar a capacidade discriminatória de um modelo individual, com os modelos mais realistas tendo uma AUC superior a 0,5. Os valores de AUC podem ser avaliados em 5 níveis de discriminação:

Valor	Classificação
[0,9;1]	Discriminação Exceccional
[0,8;0,9]	Discriminação Excelente
[0,7;0,8]	Discriminação Aceitável
[0,6;0,7]	Discriminação Fraca
[0,5;0,6]	Não há discriminação

*Tabela 2. 0.2 – Classificação dos valores de AUC*

Assim, sempre que pretendermos comparar dois modelos temos que ter em atenção o valor do AUC, e o modelo que tiver o valor de AUC mais elevado será o melhor modelo.

## ***2.5 Método de seleção de Backward (Método da seleção regressiva) <sup>9</sup>***

O método de seleção de Backward é um método que começa por incluir todas as variáveis do modelo e elimina, progressivamente, aquela cujo teste t é menos significativo, ou seja, cuja influência no ajustamento do modelo é a menor.

Após eliminar a variável menos significativa, reajusta-se o modelo com as restantes variáveis, e vai-se repetindo o processo até que todas as variáveis sejam significativas.

O modelo final depende do nível de significância  $\alpha$  escolhido para o teste t, mas é usual considerar  $\alpha=0.05$ . Este método pode produzir resultados poucos fiáveis na presença de multicolinearidade forte, uma vez que o quociente t de variáveis envolvidas em relações lineares pode conduzir a uma escolha pouco adequada da variável a retirar.

<sup>9</sup> De acordo com Alpuim, T, Modelos Lineares- Notas de apoio à disciplina, 2015.

## CAPÍTULO III – Construção do modelo

Para a construção do modelo utilizaram-se os dados das renovações de janeiro de 2016 a junho de 2017. Uma vez que o modelo começou a ser criado no 1º trimestre de 2018, apenas se considerou o 1º semestre de 2017 devido à proximidade da data de renovação. Como foi referido acima, apenas validamos a renovação com sucesso ou insucesso 50 dias depois da renovação. Assim, de forma a evitar falsos renovados considerou-se apenas os primeiros 6 meses de 2017.

Neste modelo apenas se consideram apólices de clientes individuais, em que o objeto seguro é um veículo Ligeiro, Misto ou Caminheta. O modelo não contempla apólices frota, protocolos nem ordens profissionais. Apólices saneadas (anuladas) pela empresa e anuladas antes da emissão do continuado (novo prémio para a anuidade seguinte) não são consideradas no modelo.

O tratamento dos dados antes da criação do modelo é essencial, pois proporciona resultados e conclusões mais próximas da realidade, evitando assim conclusões enviesadas.

### 3.1. *Preparação e listagem das variáveis em estudo*

A maioria dos dados trabalhados pela seguradora são fornecidos pelos clientes aquando da realização do contrato de seguro. Desta forma, existem alguns dados com erros ou falta de informação. Assim, após a construção da base de dados e da escolha das variáveis a estudo, foi necessário analisar variável a variável de forma a detetar possíveis erros, como por exemplo as idades dos condutores inferiores a 18 anos ou o caso em que as idades não se encontravam preenchidas. Nos casos em que se verificaram erros estes foram considerados *missings*.

As variáveis em estudo podem-se agrupar em 6 grupos:



Figura 3.1 – Grupos das variáveis em estudo

Assim, temos as seguintes variáveis em estudo para cada grupo:

Variável	Descrição	Tipificação das variáveis
Variáveis associadas ao cliente		
idade_tomador	Idade do tomador da apólice	Quantitativa
idade_ch	Idade do condutor do veículo seguro	Quantitativa
idade_carta	Idade de carta do condutor do veículo seguro	Quantitativa
tipo_cliente_auto	Indica se o cliente tem uma ou mais apólices automóveis	Categórica
tipo_cliente_nvida	Indica se o cliente tem um ou mais produtos não vida	Categórica
tipo_cliente_vida	Indica se o cliente tem uma ou mais apólices vida	Categórica
tipo_cliente_produto	Indica se o cliente tem um ou mais produtos não vida e vida	Categórica
Variáveis associadas à apólice		
Ano_Subscricao	Indica a data de início do contrato	Categórica
antig_apol	Indica os anos da apólice na companhia	Quantitativa
Month_Renewal	Mês em que se dá o vencimento do contrato de seguro	Quantitativa
ant_agente	Antiguidade do agente do contrato na AGEAS	Quantitativa
CDSUBRAMO	Modalidade da apólice	Categórica
CANAL_COMERCIAL	Classificação do mediador que realizou o contrato	Categórica
Formacob	Forma de cobrança	Categórica
fPagamento	Forma de fracionamento	Categórica
Capital2	Valor seguro do veículo, caso o contrato tenha as coberturas de DP	Categórica
DP	Indica se a apólice tem danos próprios	Binária
LABEL_PACK	Tipo de pack conforme as coberturas subscris	Categórica
tarifa_ini	Tarifa inicial da apólice	Categórica
agrava_5anu, agrava_3anu, agrava_1anu	Indica o número de sinistros agravantes que o cliente teve nas últimas 5, 3 e última anuidade	Quantitativa
alt_premio	Informa se houve alteração do prémio durante a anuidade anterior	Categórica
alt_bonus	Informa se houve alteração do bónus comercial na renovação	Categórica
Variáveis associadas ao veículo		
Categ	Categoria do veículo	Categórica
MarcaEmblem	Marca do veículo	Categórica
idd_veículo	Idade do veículo seguro	Quantitativa
NUM_CC2	Cilindrada do veículo	Categórica
NUM_CV2	Potência do veículo	Categórica
TARA	Tara do veículo	Categórica
CVE_COMBUSTIVEL	Combustível	Categórica
PESO_BRUTO	Peso bruto do veículo	Categórica
Valor_em_novo	Valor em novo do veículo	Categórica
Variáveis Demográficas		
UrbanRural	Classificação da morada do tomador, segundo os dados do CAMEO	Categórica
Distrito	Distrito da morada do tomador do seguro	Categórica
Variáveis associadas ao preço		
Int_Var_Rel_Depois	Intervalo de variação absoluta próxima anuidade	Categórica
Int_Var_Rel_Depois_n1	Intervalo de variação absoluta anuidade anterior	
Variáveis associadas a outros		
Cartão Ageas	Indica se o cliente utiliza ou não o cartão AGEAS	Binária

Tabela 3. 1 – Variáveis em estudo

### 3.2. Análise Descritiva

A informação disponibilizada pela companhia é extremamente confidencial, por isso, de forma a não comprometermos os dados, decidiu-se não apresentar os valores da taxa de anulação bem como a exposição. Nas secções abaixo apenas vamos poder analisar as tendências da taxa de anulação.

### 3.2.1. Variáveis associadas ao cliente

#### ❖ Idade do tomador

O gráfico abaixo apresenta-nos a relação entre a idade do tomador e a taxa de anulação. Ao observar o gráfico verificamos que a idade do tomador tem uma distribuição idêntica à normal. A partir dos 80 anos verificamos um aumento da taxa de anulação, isto deve-se provavelmente ao facto de os tomadores deixarem de poder conduzir.



Gráfico 3.1 – Taxa de anulação por Idade do tomador

#### ❖ Idade Condutor habitual

O gráfico abaixo apresenta-nos a relação entre a idade do condutor e a taxa de anulação. A partir dos 80 anos verificamos um aumento da taxa de anulação, isto deve-se provavelmente ao facto de os tomados deixarem de poder conduzir.



Gráfico 3.2 – Taxa de anulação por Idade do Condutor

### ❖ Idade Carta

O gráfico 3.3 mostra-nos a variação da taxa de anulação pela idade de carta da pessoa segura. Verificamos que os nossos segurados se encontram maioritariamente na faixa dos 20 aos 30 anos de carta. Esta variável explica-nos a experiência de condução que o segurado tem. Desta forma, quanto menor for a idade da carta menor é idade do condutor, e consequentemente maior é a taxa de anulação. Contudo, a partir dos 60 anos de carta existe pouca expressão nos dados, e por isso a taxa de anulação é bastante inconsistente.



Gráfico 3. 3 – Taxa de anulação por Idade da Carta

### ❖ Tipo de cliente automóvel

Na análise da taxa de anulação vs tipo de cliente auto verificamos que os clientes que mais anulam são aqueles que têm mais que uma apólice automóvel, possivelmente isto acontece porque caso os clientes não estejam contentes com uma das apólices cancelam todas as quem têm na companhia.

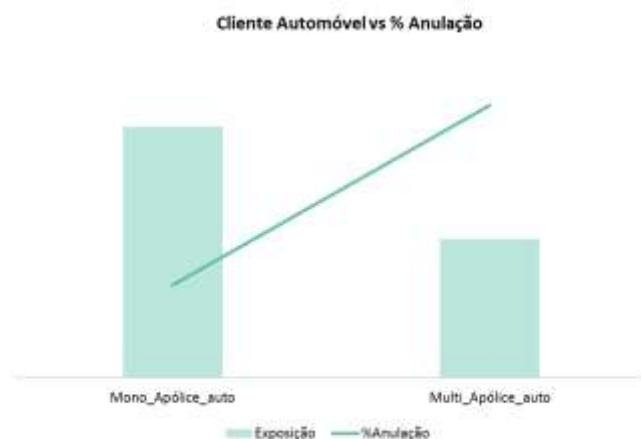


Gráfico 3. 4 - Taxa de anulação por Tipo de Cliente Automóvel

### ❖ Tipo de cliente Não-Vida

No gráfico abaixo podemos ver a evolução da taxa de anulação por tipo de cliente não vida. Como estávamos à espera, o cliente que apenas tem um produto não vida é o mais propício à anulação.

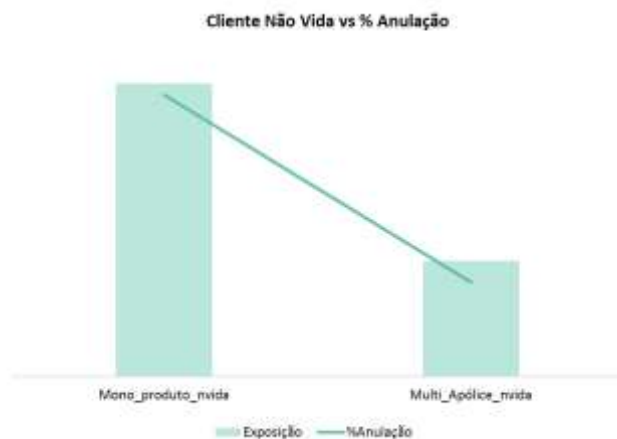


Gráfico 3. 5 - Taxa de anulação por Tipo Cliente Não-Vida

### ❖ Tipo de cliente vida

Na análise por tipo de cliente vida verificamos que os clientes que não têm produtos vida (vida risco, nesta análise não considerámos os produtos financeiros) são mais propícios à anulação. Destaca-se o facto da maioria dos nossos clientes de automóvel não terem apólices do ramo de Vida.

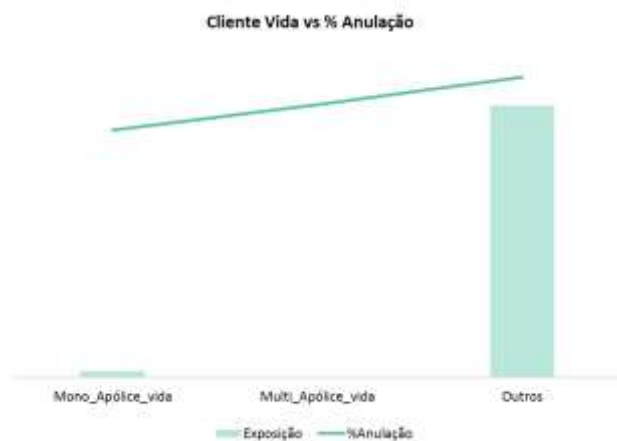


Gráfico 3. 6 - Taxa de anulação por Tipo Cliente Vida

### ❖ Tipo de cliente produto

Quando analisamos a percentagem de anulação por tipo de cliente em relação aos produtos que este tem na companhia, verificamos que os clientes que apenas têm um

produto (neste caso, apenas têm apólices automóvel) têm uma maior taxa de anulação do que os restantes. É de salientar que a maioria dos nossos clientes apenas têm um produto.

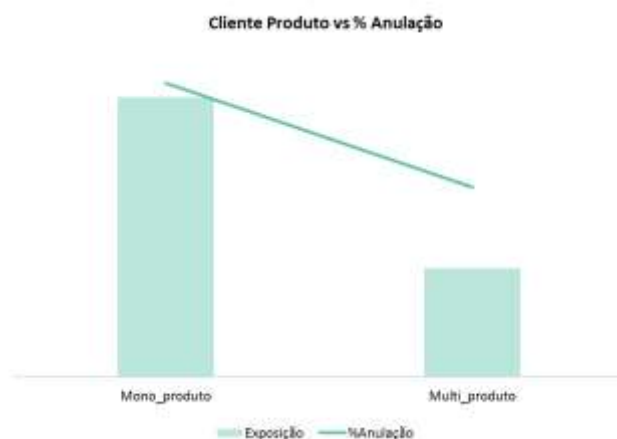


Gráfico 3. 7 – Tipo Cliente produto

### 3.2.2. Variáveis associadas à apólice

#### ❖ Ano de Subscrição

No gráfico abaixo podemos observar a tendência da taxa de anulação pelo ano de subscrição da apólice.

Rapidamente reparamos que grande parte da nossa carteira em análise concentra-se nas apólices que se iniciaram antes de 2012. No entanto o número de apólices vai aumentando até 2015. A baixa exposição em 2016 deve-se ao facto de apenas estarmos a considerar as apólices que renovam até julho de 2017 e por isso não se considera as apólices iniciadas no período julho 2016 a dezembro 2016.

Relativamente à taxa de anulação, esta é inferior nas apólices iniciadas antes de 2012. Isto deve-se ao facto de os clientes mais antigos serem menos sensíveis ao aumento do prémio. Em contrapartida, o ano de subscrição 2013 foi o que verificou uma taxa de anulação mais elevada.



Gráfico 3.8 - Taxa de anulação por Ano de Subscrição

### ❖ Antiguidade da apólice

A antiguidade da apólice na companhia é uma variável bastante importante no momento da renovação. Existem dois tipos de cliente: os que todos os anos procuram os preços mais baixos e os que têm tendência a manterem-se durante muitos anos na companhia. Como podemos verificar, a taxa de anulação é superior nas apólices que se encontram na segunda anuidade, sendo que a partir dessa anuidade a taxa de anulação diminui significativamente.



Gráfico 3. 9 - Taxa de anulação por Antiguidade da apólice

### ❖ Mês de Renovação

No gráfico abaixo podemos analisar a tendência da taxa de anulação por mês de renovação da apólice. Verificamos que as taxas mais elevadas se verificam nos meses de janeiro e agosto, este facto está relacionado com o facto de serem meses de férias e os clientes estarem mais suscetíveis a variações de prémio. No que diz respeito à exposição, tal como já foi referenciado mais acima apenas estamos a analisar o período janeiro 2016 a junho 2017, por isso é que quando analisamos por mês de renovação se verifica um decréscimo tão acentuado a partir de julho.



Gráfico 3. 10 - Taxa de anulação por Mês de renovação



### ❖ Modalidade

Como podemos verificar no gráfico abaixo, a modalidade mais vendida é a A01, os outros produtos podem ser considerados “especiais” uma vez que tem características próprias. A modalidade A17 é a que tem menor percentagem de anulação, isto deve-se ao facto deste produto manter o prémio durante 3 anos caso o cliente não tenha acidentes, e, durante esse período o valor do veículo não sofre uma desvalorização.

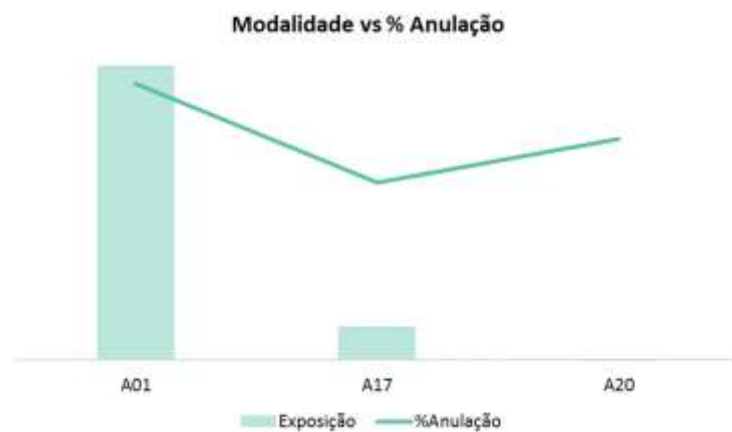


Gráfico 3.11 - Taxa de anulação por Modalidade

### ❖ Antiguidade do Agente

No gráfico seguinte podemos ver a relação entre a antiguidade do agente na AGEAS e a taxa de anulação. O agente é o intermediário entre a companhia e o segurado, e por isso a antiguidade do agente na companhia é importante no momento da renovação.

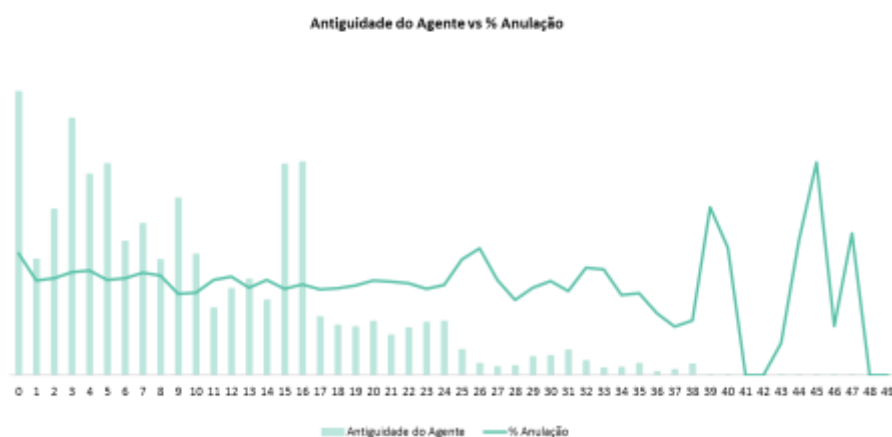


Gráfico 3.12 - Taxa de anulação por Antiguidade do Agente

### ❖ Canal Comercial

Como se pode ver através do gráfico abaixo, os principais canais de venda são os agentes exclusivos e os multimarca, sendo que a taxa de anulação mais baixa se verifica

nos agentes exclusivos. Em contrapartida, a taxa mais elevada pertencente aos agentes Outros, Corretores e Multimarca. A elevada percentagem nos agentes multimarca verifica-se porque estes agentes vendem mais do que uma marca e por isso, estão constantemente a averiguar quais os melhores preços, para proporcionar um bom serviço aos seus clientes.

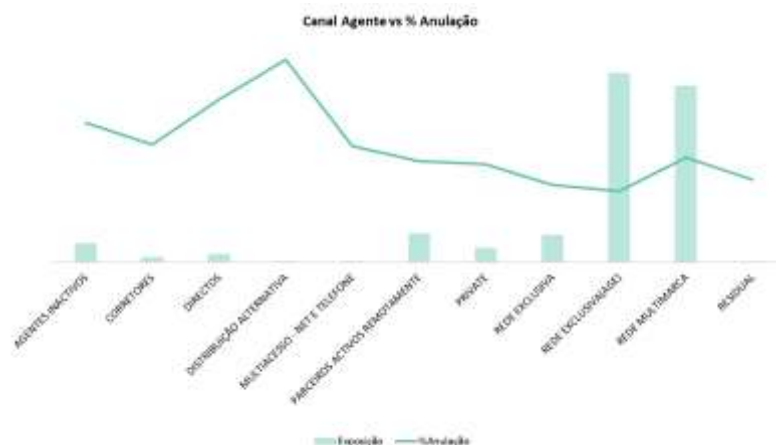


Gráfico 3.13 - Taxa de anulação por Canal do Agente

#### ❖ Forma de Cobrança

Tal como esperávamos os clientes que pagam o prémio do seguro de forma bancária (débito direto) têm uma taxa de anulação inferior aos que não pagam dessa forma. Isto verifica-se porque os clientes que pagam de forma bancária têm tendência a não tomar tanta atenção ao valor que lhes é cobrado na renovação.

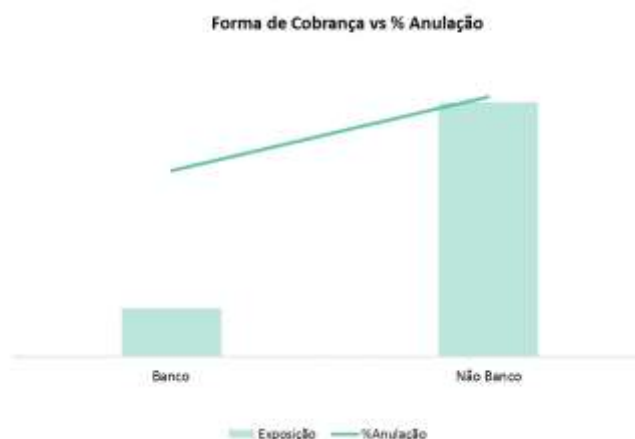


Gráfico 3.14 - Taxa de anulação por Forma de Cobrança

#### ❖ Forma de Pagamento

No gráfico abaixo podemos ver o comportamento da taxa de anulação de acordo com a forma de pagamento. Tal como podemos verificar, um cliente que paga o seu

prémio de forma anual anula muito mais do que aquele que paga mensalmente. Isto acontece devido ao facto de quem paga anualmente sente mais o aumento do que aquele que paga mensalmente. No entanto, a grande parte da nossa carteira diz respeito a clientes que pagam anualmente.

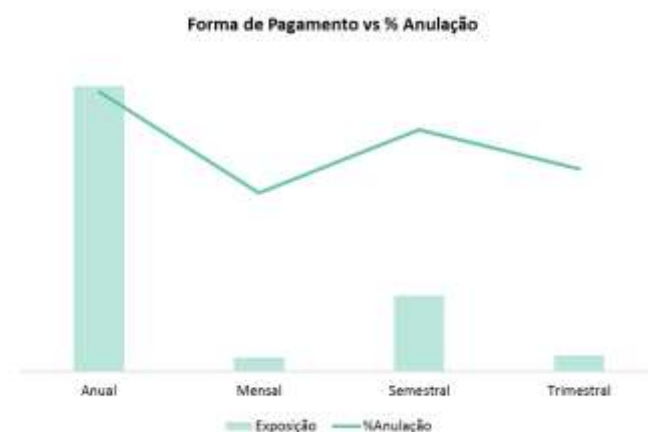


Gráfico 3.15 - Taxa de anulação por Forma de Pagamento

#### ❖ Capital da apólice

O Gráfico 3.16 mostra como varia a taxa de anulação consoante o capital da apólice. O capital da apólice diz respeito ao valor seguro do veículo caso este tenha seguro de danos próprios. Quando um veículo não apresenta informação do capital é porque apenas tem seguro de responsabilidade civil. De forma a ser mais fácil interpretar esta análise, o gráfico 3.17 apenas nos mostra as apólices com danos próprios. Pela análise ao gráfico verificamos que os veículos com valor seguro mais elevado são mais propícios a anular.

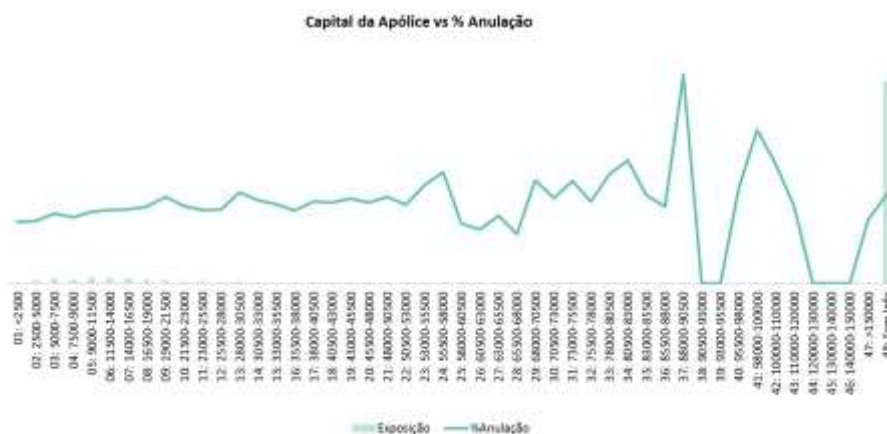


Gráfico 3.16 - Taxa de anulação por Capital da Apólice (1)

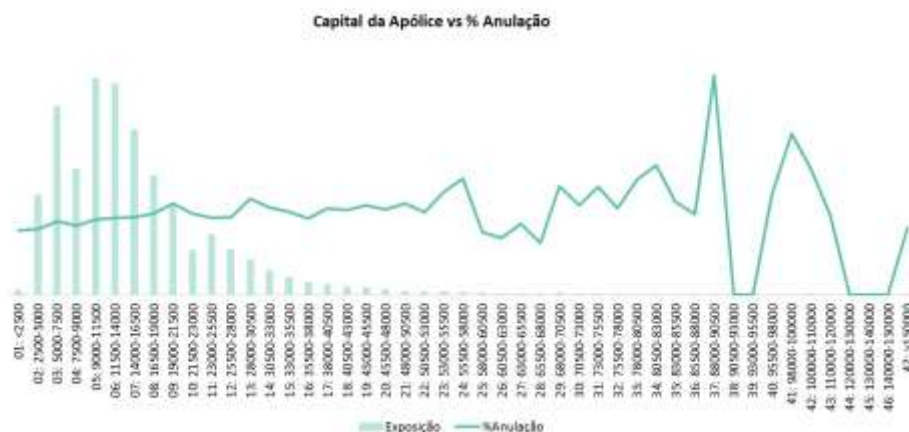


Gráfico 3.17 - Taxa de anulação por Capital da Apólice (2)

#### ❖ DP

No gráfico subsequente podemos ver a relação da taxa de anulação com o tipo de pack da apólice, dividido por Responsabilidade Civil (obrigatório) e Danos Próprios. Verificamos que a taxa de anulação é superior nos contratos que apenas têm a cobertura obrigatória, Responsabilidade Civil.



Gráfico 3.18 - Taxa de anulação por Tipo de Pack

#### ❖ Pack da apólice

Quando analisamos a taxa de anulação por pack da apólice, estando este mais segmentado do que o do gráfico acima verificamos que o pack com maior taxa de anulação é o que apenas tem a cobertura de responsabilidade civil. O pack com menor taxa de anulação é o que têm as coberturas de Furto ou Roubo e Incêndio, Raio ou Explosão. No entanto, podemos verificar que o pack com maior número de apólices é o que inclui as coberturas de Responsabilidade Civil e de QIV (Queda Isolada de Vidros).

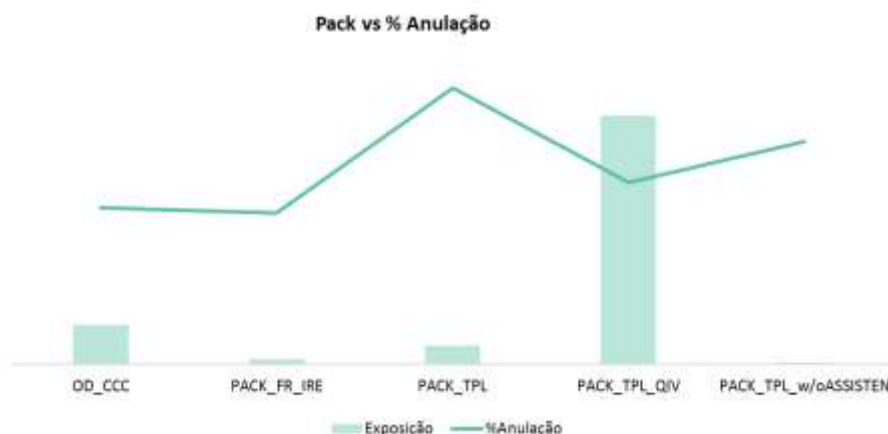


Gráfico 3.19 - Taxa de anulação por Pack da Apólice

### ❖ Tarifa inicial

É essencial às seguradoras irem atualizado as tarifas que estão a ser comercializadas, pois ano após ano verifica-se o aumento da competitividade no mercado segurador em Portugal, e é essencial ter um preço ajustado. No gráfico abaixo podemos ver a relação entre a tarifa inicial e a taxa de anulação. Verificamos que as tarifas 9C e 9D são as que têm maior expressão, uma vez que são as tarifas mais recentes nos dados utilizados. Nestas tarifas a taxa de anulação é idêntica. Como era expectável, a taxa de anulação é superior nas tarifas que tem menor expressão.

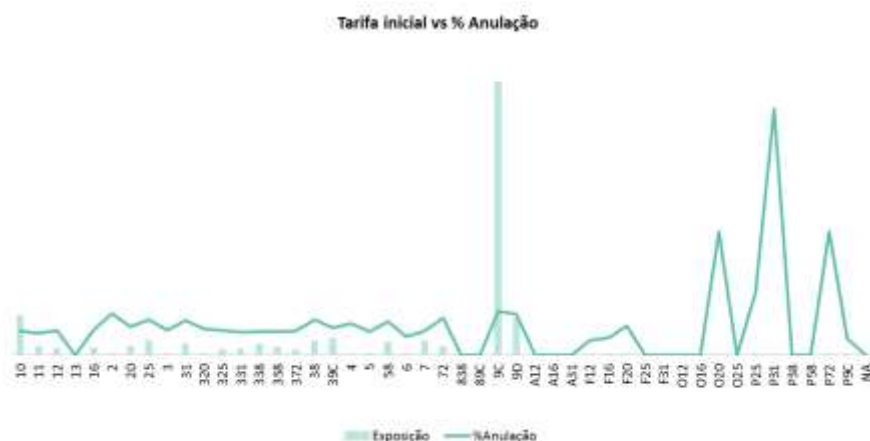


Gráfico 3.20 - Taxa de anulação por Tarifa inicial da Apólice

### ❖ Sinistros agravantes nas últimas 5 anuidades, 3 anuidades e na última anuidade

Como era expectável, a taxa de anulação é superior nas apólices que tiveram sinistros agravantes, que são sinistros que ocorreram nas coberturas: Responsabilidade Civil (obrigatória e facultativa), Choque, Colisão ou Capotamento e Incêndio, Raio ou Explosão, Furto ou Roubo, Só colisão e solução ano seguro, nas últimas 5 e 3 anuidades.

Entenda-se por sinistros agravantes aqueles que agravam o preço da apólice na anuidade seguinte ao sinistro. O mesmo se verifica quando analisamos os sinistros agravantes da última anuidade.

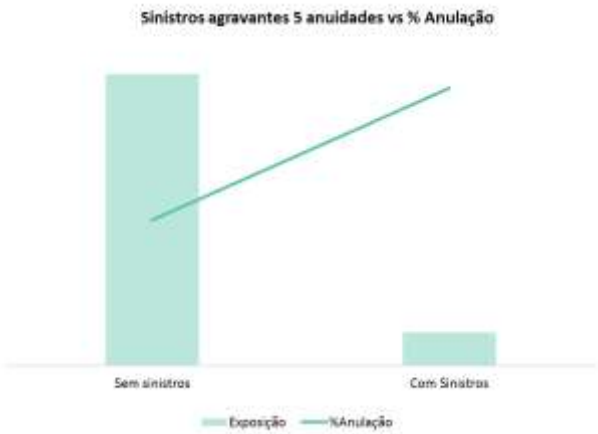


Gráfico 3. 21 – Taxa de anulação por sinistros agravantes nas últimas 5 anuidades

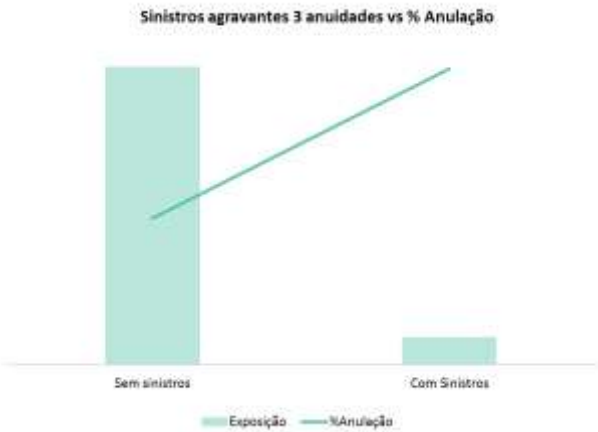


Gráfico 3. 22 – Taxa de anulação por sinistros agravantes nas últimas 3 anuidades



Gráfico 3.23 - Taxa de anulação por sinistros agravantes na última anuidade

### ❖ Alteração do prémio

No gráfico abaixo verificamos a reação do cliente face às alterações do prémio durante a anuidade. Estas alterações de prémio podem-se verificar, devido, por exemplo, a alterações de veículo, coberturas, entre outras. Deste modo, verificamos que a taxa de anulação é superior nos casos em que o prémio aumenta. Contudo na maioria das apólices o prémio não se altera.



Gráfico 3. 24 - Taxa de anulação por Alteração do prémio

### ❖ Alteração do Bónus

No gráfico seguinte podemos analisar a taxa de anulação de acordo com a alteração do Bónus na renovação. Existem duas situações que verificamos imediatamente: a maioria das apólices aumenta o escalão bónus na renovação, isto significa que não existiram sinistros na anuidade que termina; e a taxa de anulação é superior nos casos em que o bónus diminuiu. O bónus diminuiu quando existe um sinistro, desta forma, o prémio é agravado duas vezes, uma por diminuição do bónus e outra por aumento tarifário devido à existência de sinistro.

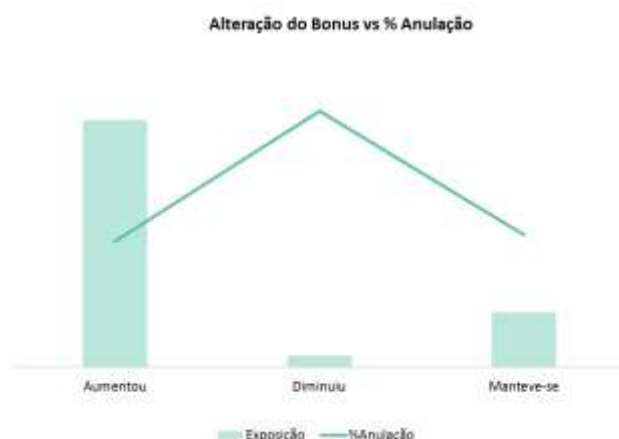


Gráfico 3. 25 - Taxa de anulação por Alteração do escalão de bónus





### ❖ Idade do veículo

No gráfico seguinte podemos observar a variação da taxa de anulação pela idade do veículo. É possível verificar que até aos 30 anos a taxa de anulação mantém-se constante, verificando-se um ligeiro aumento. Isto acontece, pois, os clientes tendem a adquirir novos veículos e consequentemente a anular a apólice anterior. Como a partir dos 30 anos a exposição é muito baixa, isto leva a que a taxa de anulação seja oscilante.



Gráfico 3. 28 - Taxa de anulação por Idade do veículo

### ❖ NUM\_CC

O gráfico 3.29 mostra-nos a variação da taxa de anulação pela cilindragem do veículo seguro. A maioria da carteira encontra-se em duas bandas: entre os 1100 e 1200 e entre 1300 e os 1400. Através da observação do gráfico podemos verificar que a taxa de anulação não é consistente ao longo dos vários intervalos.

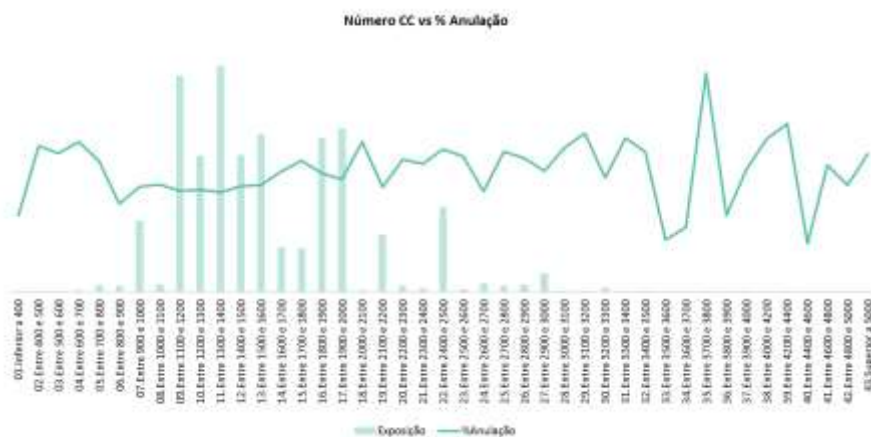


Gráfico 3.29 - Taxa de anulação por número de CC

## ❖ NUM\_CV

O gráfico 3.30 mostra-nos a variação da taxa de anulação pela potência do veículo seguro, a maioria da carteira encontra-se em duas bandas: entre os 60 e 70 e entre 70 e os 80. Podemos verificar que a taxa de anulação se mantém praticamente constante até ao intervalo “19. Entre 210 e 220”. No entanto a exposição também é baixa, pelo que a taxa de anulação é volátil.

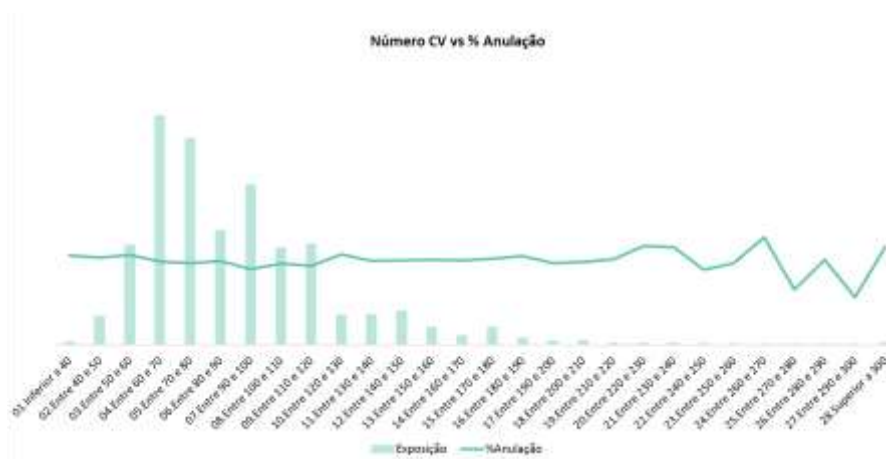


Gráfico 3. 30 - Taxa de anulação por número de CV

## ❖ TARA

O gráfico 3.31 mostra-nos a variação da taxa de anulação pela tara do veículo seguro. A Tara é o peso do veículo, sem passageiros nem carga. Através do gráfico abaixo verificamos que a taxa de anulação é inferior nas bandas onde se encontra maior exposição de veículos, começando a aumentar a partir da banda “27. Entre 1850 e 1900”.

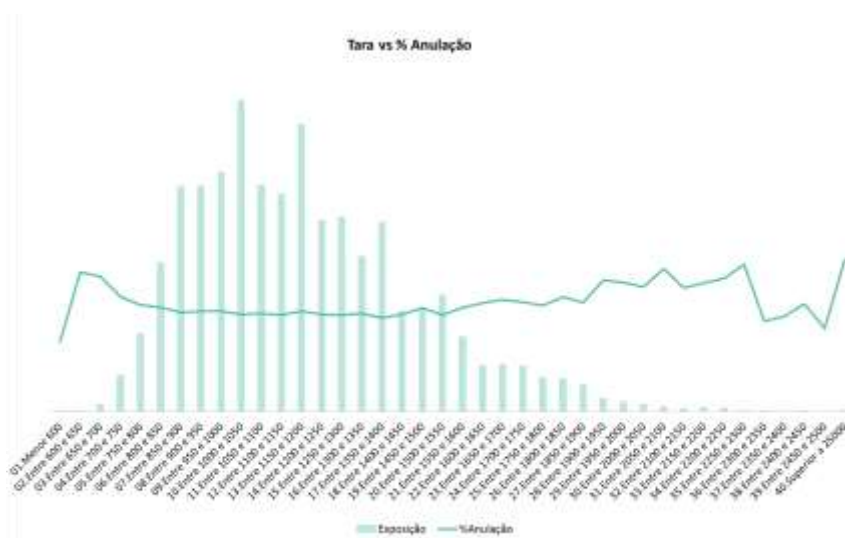


Gráfico 3.31 - Taxa de anulação por Tara do veículo

## ❖ Combustível

Relativamente ao gráfico 3.32 podemos verificar que a taxa de anulação relativa aos veículos a Diesel é substancialmente superior à dos veículos a Gasolina. Dentro do grupo “Outros” estamos a considerar os veículos elétricos, híbridos e mistos. Como a exposição deste tipo de veículos é insignificante isto leva a que a taxa de anulação seja mais elevada. No entanto, este tipo de veículos estão a ganhar cada vez mais representatividade no mercado.

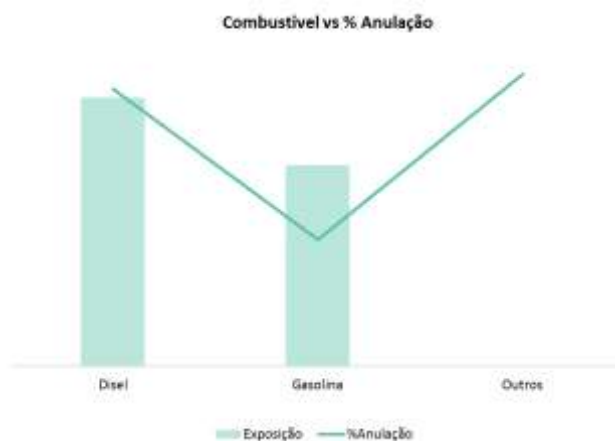


Gráfico 3.32 - Taxa de anulação por Tipo de Combustível

## ❖ Peso Bruto

O gráfico abaixo apresenta-nos a relação entre Peso Bruto do veículo e a taxa de anulação. Como podemos verificar, esta variável não é consistente ao longo das bandas. E existe uma grande parte dos dados em análise que não têm informação acerca desta variável.

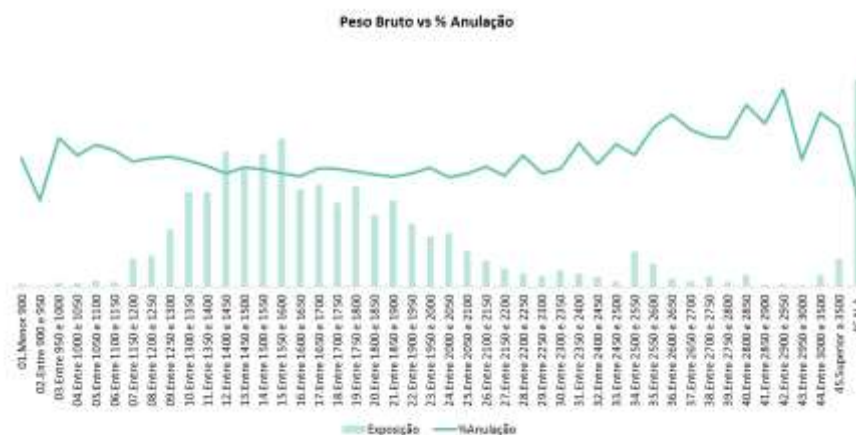


Gráfico 3.33 - Taxa de anulação por Peso Bruto do veículo

### ❖ Valor em novo

O gráfico 3.34 mostra-nos a relação entre o valor do veículo em novo e a taxa de anulação. É expectável que os tomadores que tenham veículos com valor mais baixo sejam mais sensíveis às variações do prémio. A maioria dos veículos do modelo são veículos com valor em novo entre o 8000€ e os 25000€. É de realçar o facto dos veículos mais dispendiosos terem taxas de anulação mais elevadas e mais instáveis.



Gráfico 3.34 - Taxa de anulação por Valor em novo do veículo

### 3.2.4. Variáveis Demográficas

#### ❖ Urban – Rural

A variável Urban-Rural, apresentada no gráfico abaixo, é uma variável externa, esta variável classifica as localidades como Urbanas ou Rurais de acordo com o seu código postal. De acordo com o gráfico abaixo verificamos que os clientes que habitam em zonas rurais são mais propícios ao cancelamento, sendo estes os que estão em maioria na nossa carteira.

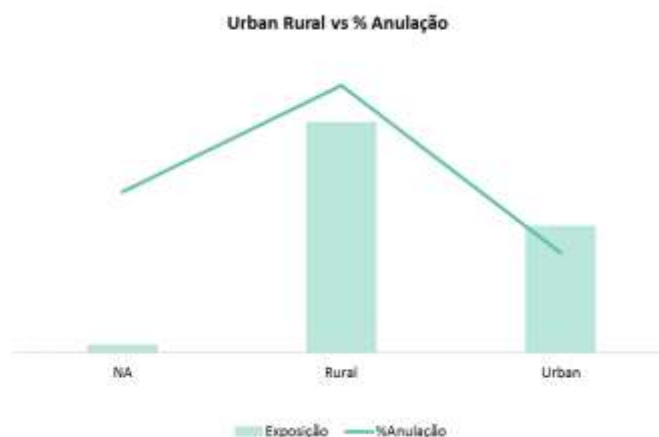


Gráfico 3.35 - Taxa de anulação por Urban Rural

## ❖ Distrito

Através da análise ao gráfico 3.36 podemos verificar que grande parte da nossa carteira encontra-se no Porto e em Aveiro, devido à própria história da companhia, seguindo-se de Lisboa. Analisando a taxa de anulação verificamos que os Açores é o distrito com maior taxa de anulação, no entanto existe pouca exposição nesse distrito. De forma a ser possível analisar os restantes distritos, retirou-se os Açores e o Estrangeiro. No gráfico 3.37 verificamos que a taxa de anulação é superior em Portalegre.

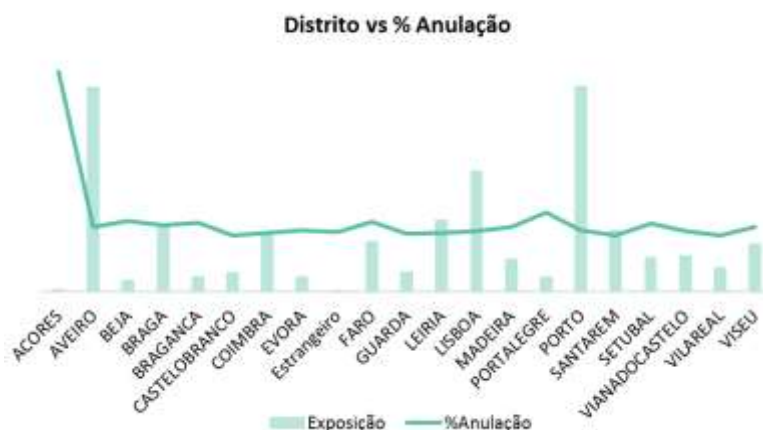


Gráfico 3.36 - Taxa de anulação por Distrito (1)



Gráfico 3.37 - Taxa de anulação por Distrito (2)

### 3.2.5. Variáveis associadas ao preço

#### ❖ Intervalo de Variação Relativa

O gráfico abaixo apresenta-nos a relação entre o intervalo de variação relativa e a taxa de anulação. Esta variação é a verificada entre o prémio da anuidade anterior à renovação e a posterior. Podemos verificar que a maioria das apólices sofre um aumento entre os 0% e 0,5% e entre 4% e 5%. Como era expectável, quanto maior é o aumento, mais elevada é a taxa de anulação.

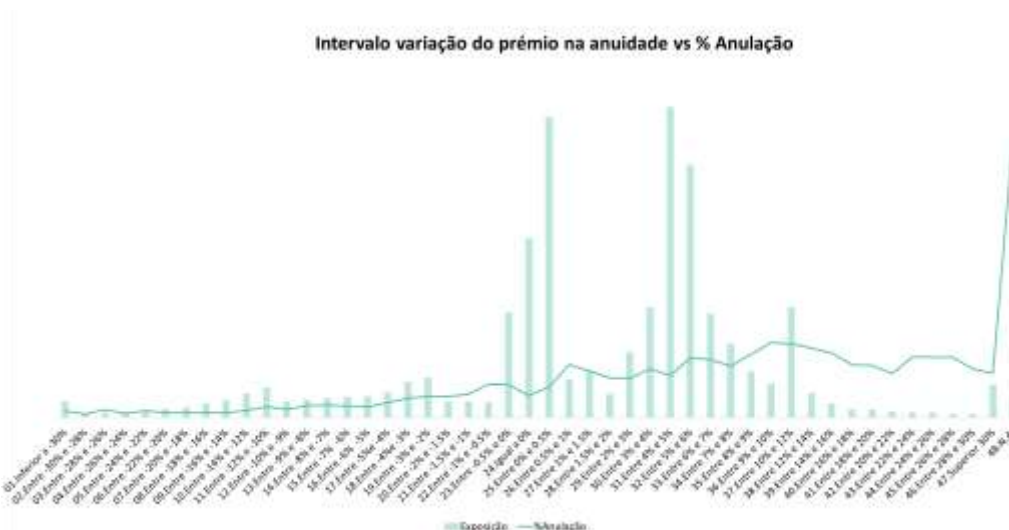


Gráfico 3.38 - Taxa de anulação por Intervalo de Variação Relativa

#### ❖ Intervalo de Variação Relativa anuidade anterior

O gráfico 3.39 diferencia-se do gráfico anterior uma vez que neste verificamos a variação do prêmio na anuidade anterior. Tal como no gráfico 3.38, quanto maior é o aumento, mais elevada é a taxa de anulação. No entanto, na anuidade anterior verifica-se que houve mais apólices a sofrerem um aumento entre os 7% e 8%.



Gráfico 3. 39 - Taxa de anulação por Intervalo de Variação Relativa na anuidade anterior

### 3.2.6. Variáveis associadas a outros

#### ❖ Cartão Mundo AGEAS

O cartão mundo AGEAS é um cartão a que todos os clientes têm acesso. Este cartão dá direito a descontos nas lojas com protocolo com a Ageas. Podemos verificar

que a maioria dos clientes que possuem o cartão não o utiliza, e simultaneamente a taxa de anulação é mais elevada nesses casos.

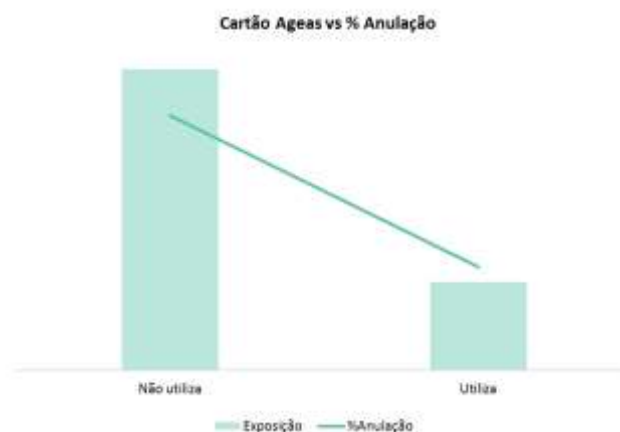


Gráfico 3. 40 - Taxa de anulação por Cartão Ageas

### 3.3. Modelização

Para modelar a base de dados criada no software SAS Guide, importou-se a mesma para o software Emblem. Este software é bastante utilizado nas equipas de *Pricing*, uma vez que permite criar modelos preditivos complexos.

#### 3.3.1. Correlações

A base criada para modelar continha muitas variáveis e algumas delas correlacionadas entre si. Desta forma, para o modelo estar perfeitamente adequado, as variáveis altamente correlacionadas foram removidas do modelo. Para avaliar as correlações utilizou-se o V-Cramer, utilizando um limite de 0,7. No quadro abaixo podemos ver a matriz de correlações V-Cramer já reduzida apenas com as variáveis correlacionadas:

Variável I	Variável II	V-Cramer
Label Pack	DP	1
tipo_cliente_nvida	Tipo_cliente_auto	0.754
Capital	DP	1
Agrava_5anu	Agrava_3anu	0.72
Idade condutor	Idade tomador	0.926

Tabela 3. 2 – Variáveis altamente correlacionadas

As variáveis assinaladas a cinzento na tabela foram retiradas do modelo. A escolha das variáveis a manter prende-se com os seguintes motivos:

- São mais relevantes, do ponto de vista do negócio;
- Os dados apresentam uma qualidade superior;

- Variáveis com maior granularidade.

### 3.3.2. Métodos de seleção de variáveis

O passo seguinte ao retirar as variáveis correlacionadas do modelo será utilizar os métodos de seleção de variáveis para a construção do modelo final. Utilizou-se o método de *Backwards* e o Teste de *Wald*, sendo que o critério de escolha foi o que resultasse no menor AIC

#### ❖ Método de *Backward*

O método de *Backward* inicia-se com todas as variáveis do modelo. Eliminou-se, progressivamente, aquelas com *p-value* superior a 5%, uma vez que se tomou  $\alpha=0.05$ .

#### 1º Iteração:

A variável “Tara” é a variável com o maior *p-value*, superior a 5%, desta forma esta é a primeira a ser retirada do modelo.

Variável	AIC	F	Variável	AIC	F
Current Model Stat	197453.50	0%	tipo_cliente_vida	197445.84	0%
Ano_Subscricao	199187.11	0%	tipo_cliente_produto	197783.10	0%
Distrito_apol	197808.01	0%	idade_tomador	197394.62	4%
CANAL_AGENTE	198609.13	0%	agrava_5anu	197701.04	0%
Month_Renewal	197546.14	0%	agrava_1anu	197460.10	0%
UrbanRural	197448.46	20%	cartao	197530.79	0%
CVE_TARIFA	198997.27	0%	capital2	197408.48	24%
formacob	197634.52	0%	idade_carta	197435.16	0%
LABEL_PACK	197806.78	0%	antig_apol	197496.03	0%
CVE_COMBUSTIVEL	197453.41	9%	fPagamento	198196.87	0%
categ	197913.25	0%	Int_Var_Rel_Depois	231145.93	0%
MarcaEmblem	197761.73	0%	Int_Var_Rel_Depois_n1	201855.56	0%
alt_premio	197848.81	0%	NUM_CC2	197420.73	12%
alt_bonus	200072.66	0%	NUM_CV2	197452.57	0%
idd_veiculo	197600.04	0%	TARA	197400.6701	63%
ant_agente	197610.06	0%	PESO_BRUTO	197415.17	10%
tipo_cliente_auto	197516.03	0%	Valor_em_novo	197361.62	57%

Tabela 3.3 – 1ª Iteração do método de *Backwards*



## 2º Iteração:

A variável “tipo\_cliente\_vida” é a variável com o maior *p-value*, superior a 5%, desta forma esta é a segunda a ser retirada do modelo.

Variável	AIC	F	Variável	AIC	F
Current Model Stat	197453.50	0%	tipo_cliente_vida	197421.50	93%
Ano_Subscricao	199188.76	0%	tipo_cliente_produto	197772.35	0%
Distrito_apol	197827.96	0%	idade_tomador	197377.89	10%
CANAL_AGENTE	198655.61	0%	agrava_5anu	197991.13	0%
Month_Renewal	197570.68	0%	agrava_1anu	197992.27	0%
UrbanRural	197439.41	46%	cartao	198033.26	0%
CVE_TARIFA	199028.97	0%	capital2	197410.35	32%
formacob	197653.52	0%	idade_carta	197433.99	0%
LABEL_PACK	197837.12	0%	antig_apol	197502.81	0%
CVE_COMBUSTIVEL	197463.39	0%	fPagamento	198198.19	0%
categ	197901.01	0%	Int_Var_Rel_Depois	231130.41	0%
MarcaEmblem	197668.03	0%	Int_Var_Rel_Depois_n1	201856.59	0%
alt_premio	197560.72	0%	NUM_CC2	197429.40	4%
alt_bonus	200042.12	0%	NUM_CV2	197469.96	0%
idd_veiculo	197652.54	0%	PESO_BRUTO	197418.39	11%
ant_agente	197616.56	0%	Valor_em_novo	197385.00	37%
tipo_cliente_auto	197506.77	0%			

Tabela 3.4 - 2ª Iteração do método de Backwards

## 3º Iteração:

A variável “UrbanRural” é a variável com o maior *p-value*, superior a 5%, desta forma esta é a terceira a ser retirada do modelo.

Variável	AIC	F	Variável	AIC	F
Current Model Stat	197445.68	0%	tipo_cliente_auto	197510.54	0%
Ano_Subscricao	199207.84	0%	tipo_cliente_produto	197776.02	0%
Distrito_apol	197860.14	0%	idade_tomador	197389.92	2%
CANAL_AGENTE	198660.90	0%	agrava_5anu	197673.56	0%
Month_Renewal	197547.36	0%	agrava_1anu	197493.82	0%
UrbanRural	197422.74	99%	cartao	197537.04	0%
CVE_TARIFA	198991.68	0%	capital2	197402.00	28%
formacob	197650.52	0%	idade_carta	197411.36	0%
LABEL_PACK	197817.10	0%	antig_apol	197477.72	0%
CVE_COMBUSTIVEL	197433.04	0%	fPagamento	198208.36	0%
categ	197899.43	0%	Int_Var_Rel_Depois	231130.89	0%
MarcaEmblem	197462.79	0%	Int_Var_Rel_Depois_n1	201857.74	0%
alt_premio	197561.08	0%	NUM_CC2	197412.36	4%
alt_bonus	200076.79	0%	NUM_CV2	197469.76	0%
idd_veiculo	197637.46	0%	PESO_BRUTO	197420.09	6%
ant_agente	197619.64	0%	Valor_em_novo	197379.46	21%

Tabela 3.5 - 3ª Iteração do método de Backwards

#### **4º Iteração:**

A variável “capital2” é a variável com o maior *p-value*, superior a 5%, desta forma esta é a quarta a ser retirada do modelo.

Variável	AIC	F	Variável	AIC	F
Current Model Stat	197447.66	0%	tipo_cliente_produto	198189.58	0%
Ano_Subscricao	199207.10	0%	idade_tomador	197382.87	11%
Distrito_apol	197860.89	0%	agrava_5anu	197648.08	0%
CANAL_AGENTE	198671.40	0%	agrava_1anu	197497.96	0%
Month_Renewal	197544.35	0%	cartao	197540.48	0%
CVE_TARIFA	198990.24	0%	capital2	197401.37	57%
formacob	197668.48	0%	idade_carta	197430.68	0%
LABEL_PACK	197830.58	0%	antig_apol	197481.30	0%
CVE_COMBUSTIVEL	197453.03	0%	fPagamento	198170.94	0%
categ	197908.48	0%	Int_Var_Rel_Depois	231130.76	0%
MarcaEmblem	197439.44	0%	Int_Var_Rel_Depois_n1	201865.00	0%
alt_premio	197590.62	0%	NUM_CC2	197396.44	23%
alt_bonus	200074.24	0%	NUM_CV2	197469.76	0%
idd_veiculo	197604.68	0%	PESO_BRUTO	197432.27	1%
ant_agente	197604.39	0%	Valor_em_novo	197382.70	24%
tipo_cliente_auto	197630.47	0%			

*Tabela 3.6 - 4ª Iteração do método de Backwards*

#### **5º Iteração:**

A variável “Valor\_em\_novo” é a variável com o maior *p-value*, superior a 5%, desta forma esta é a quinta a ser retirada do modelo.

Variável	AIC	F	Variável	AIC	F
Current Model Stat	197442.91	0%	tipo_cliente_auto	197624.51	0%
Ano_Subscricao	199199.52	0%	tipo_cliente_produto	198134.04	0%
Distrito_apol	197857.22	0%	idade_tomador	197371.16	10%
CANAL_AGENTE	198647.92	0%	agrava_5anu	197671.24	0%
Month_Renewal	197525.58	0%	agrava_1anu	197497.16	0%
CVE_TARIFA	199005.78	0%	cartao	197539.84	0%
formacob	197639.34	0%	idade_carta	197429.29	0%
LABEL_PACK	197834.50	0%	antig_apol	197498.52	0%
CVE_COMBUSTIVEL	197444.54	4%	fPagamento	198174.88	0%
categ	197880.49	0%	Int_Var_Rel_Depois	231133.76	0%
MarcaEmblem	197462.45	0%	Int_Var_Rel_Depois_n1	201864.20	0%
alt_premio	197561.02	0%	NUM_CC2	197393.73	22%
alt_bonus	200075.36	0%	NUM_CV2	197444.45	0%
idd_veiculo	197618.44	0%	PESO_BRUTO	197421.25	4%
ant_agente	197614.25	0%	Valor_em_novo	197384.55	27%

*Tabela 3.7 - 5ª Iteração do método de Backwards*

### **6º Iteração:**

A variável “PESO\_BRUTO” é a variável com o maior *p-value*, superior a 5%, desta forma esta é a sexta a ser retirada do modelo.

Variável	AIC	F	Variável	AIC	F
Current Model Stat	197400.96	0%	tipo_cliente_auto	197585.23	0%
Ano_Subscricao	199147.04	0%	tipo_cliente_produto	198157.93	0%
Distrito_apol	197816.13	0%	idade_tomador	197354.53	1%
CANAL_AGENTE	198613.35	0%	agrava_5anu	197619.38	0%
Month_Renewal	197497.70	0%	agrava_1anu	197455.78	0%
CVE_TARIFA	198935.53	0%	cartao	197493.01	0%
formacob	197606.64	0%	idade_carta	197384.24	0%
LABEL_PACK	197790.47	0%	antig_apol	197432.65	0%
CVE_COMBUSTIVEL	197407.48	0%	fPagamento	198129.12	0%
categ	197887.59	0%	Int_Var_Rel_Depois	231094.09	0%
MarcaEmblem	208633.05	0%	Int_Var_Rel_Depois_n1	201826.78	0%
alt_premio	208766.83	0%	NUM_CC2	197382.29	2%
alt_bonus	200033.31	0%	NUM_CV2	197420.98	0%
idd_veiculo	197575.37	0%	PESO_BRUTO	197350.97	24%
ant_agente	197563.95	0%			

*Tabela 3. 8 - 6ª Iteração do método de Backwards*

### **7º Iteração:**

A variável “CVE\_COMBUSTIVEL” é a variável com o maior *p-value*, superior a 5%, desta forma esta é a sétima a ser retirada do modelo.

Variável	AIC	F	Variável	AIC	F
Current Model Stat	197359.88	0%	ant_agente	197538.04	0%
Ano_Subscricao	199098.59	0%	tipo_cliente_auto	197544.60	0%
Distrito_apol	197749.28	0%	tipo_cliente_produto	198094.29	0%
CANAL_AGENTE	198552.73	0%	idade_tomador	197300.95	4%
Month_Renewal	197464.81	0%	agrava_5anu	197566.89	0%
CVE_TARIFA	198910.23	0%	agrava_1anu	197381.30	0%
formacob	197552.24	0%	cartao	197449.95	0%
LABEL_PACK	197732.42	0%	idade_carta	197347.65	0%
CVE_COMBUSTIVEL	197351.91	13%	antig_apol	197426.19	0%
categ	197838.05	0%	fPagamento	198107.51	0%
MarcaEmblem	197371.91	0%	Int_Var_Rel_Depois	231031.43	0%
alt_premio	197499.62	0%	Int_Var_Rel_Depois_n1	201784.78	0%
alt_bonus	199989.01	0%	NUM_CC2	197348.30	0%
idd_veiculo	197527.65	0%	NUM_CV2	197372.30	0%

*Tabela 3.9 - 7ª Iteração do método de Backwards*

### 8º Iteração:

Todas as variáveis do modelo têm um valor de *p-value* inferior a 5%. Desta forma, não retiramos mais nenhuma variável ao modelo, e assim o método de *Backwards* termina com 27 variáveis. Sendo o AIC=197360.8.

Variável	AIC	F	Variável	AIC	F
Current Model Stat	197360.78	0%	tipo_cliente_auto	197548.43	0%
Ano_Subscricao	199108.31	0%	tipo_cliente_produto	198095.88	0%
Distrito_apol	197778.18	0%	idade_tomador	197301.39	5%
CANAL_AGENTE	198566.36	0%	agrava_5anu	197573.39	0%
Month_Renewal	197464.34	0%	agrava_1anu	197384.57	0%
CVE_TARIFA	198910.80	0%	cartao	197426.23	0%
formacob	197563.33	0%	idade_carta	197332.80	0%
LABEL_PACK	197745.29	0%	antig_apol	197421.77	0%
categ	197831.52	0%	fPagamento	198110.18	0%
MarcaEmblem	197367.41	0%	Int_Var_Rel_Depois	231040.52	0%
alt_premio	197482.20	0%	Int_Var_Rel_Depois_n1	201780.50	0%
alt_bonus	199990.94	0%	NUM_CC2	197361.00	0%
idd_veiculo	197582.17	0%	NUM_CV2	197406.46	0%
ant_agente	197544.94	0%			

Tabela 3.10 - 8ª Iteração do método de Backwards

### ❖ Teste de Wald

No teste de *Wald* iniciamos com todas as variáveis e vamos retirando uma a uma, de acordo com o valor de *p-value*. Retira-se as variáveis com *p-value* mais elevado em cada uma das iterações.

### 1º Iteração:

A variável “capital2” é a variável com o maior *p-value*, superior a 5%, desta forma esta é a primeira a ser retirada do modelo.

Variável	Wald p Value (Chi-Sq)	Variável	Wald p Value (Chi-Sq)
capital2	94%	formacob	0%
CVE_COMBUSTIVEL	8%	LABEL_PACK	0%
TARA	62%	categ	0%
NUM_CC2	4%	alt_premio	0%
PESO_BRUTO	3%	alt_bonus	0%
Valor_em_novo	26%	idd_veiculo	0%
idade_tomador	3%	ant_agente	0%
UrbanRural	20%	tipo_cliente_auto	0%
tipo_cliente_nvida	17%	tipo_cliente_produto	0%
tipo_cliente_vida	14%	agrava_5anu	0%
idade_carta	0%	agrava_1anu	0%
NUM_CV2	0%	cartao	0%
MarcaEmblem	0%	antig_apol	0%
Ano_Subscricao	0%	fPagamento	0%
Distrito_apol	0%	Int_Var_Rel_Depois	0%
CANAL_AGENTE	0%	Int_Var_Rel_Depois_n1	0%
Month_Renewal	0%	CDSUBRAMO	0%
CVE_TARIFA	0%		

Tabela 3. 11 - 1ª Iteração do teste de Wald

## **2º Iteração:**

A variável “TARA” é a variável com o maior *p-value*, superior a 5%, desta forma esta é a segunda a ser retirada do modelo.

Variável	Wald p Value (Chi-Sq)	Variável	Wald p Value (Chi-Sq)
TARA	56%	formacob	0%
PESO_BRUTO	4%	LABEL_PACK	0%
Valor_em_novo	24%	categ	0%
idade_tomador	3%	alt_premio	0%
NUM_CC2	3%	alt_bonus	0%
UrbanRural	20%	idd_veiculo	0%
tipo_cliente_nvida	16%	ant_agente	0%
tipo_cliente_vida	13%	tipo_cliente_auto	0%
CVE_COMBUSTIVEL	11%	tipo_cliente_produto	0%
idade_carta	0%	agrava_5anu	0%
MarcaEmblem	0%	agrava_1anu	0%
NUM_CV2	0%	cartao	0%
Ano_Subscricao	0%	antig_apol	0%
Distrito_apol	0%	fPagamento	0%
CANAL_AGENTE	0%	Int_Var_Rel_Depois	0%
Month_Renewal	0%	Int_Var_Rel_Depois_n1	0%
CVE_TARIFA	0%	CDSUBRAMO	0%

*Tabela 3.12 - 2ª Iteração do teste de Wald*

## **3º Iteração:**

A variável “UrbanRural” é a variável com o maior *p-value*, superior a 5%, desta forma esta é a terceira a ser retirada do modelo.

Variável	Wald p Value (Chi-Sq)	Variável	Wald p Value (Chi-Sq)
CVE_COMBUSTIVEL	7%	LABEL_PACK	0%
idade_tomador	3%	categ	0%
UrbanRural	21%	alt_premio	0%
PESO_BRUTO	3%	alt_bonus	0%
Valor_em_novo	20%	idd_veiculo	0%
tipo_cliente_nvida	17%	ant_agente	0%
tipo_cliente_vida	13%	tipo_cliente_auto	0%
NUM_CC2	1%	tipo_cliente_produto	0%
idade_carta	0%	agrava_5anu	0%
MarcaEmblem	0%	agrava_1anu	0%
NUM_CV2	0%	cartao	0%
Ano_Subscricao	0%	antig_apol	0%
Distrito_apol	0%	fPagamento	0%
CANAL_AGENTE	0%	Int_Var_Rel_Depois	0%
Month_Renewal	0%	Int_Var_Rel_Depois_n1	0%
CVE_TARIFA	0%	CDSUBRAMO	0%
formacob	0%		

*Tabela 3.13 - 2ª Iteração do teste de Wald*

#### **4º Iteração:**

A variável “Valor\_em\_novo” é a variável com o maior *p-value*, superior a 5%, desta forma esta é a quarta a ser retirada do modelo.

Variável	Wald p Value (Chi-Sq)	Variável	Wald p Value (Chi-Sq)
CVE_COMBUSTIVEL	6%	LABEL_PACK	0%
PESO_BRUTO	3%	categ	0%
idade_tomador	3%	alt_premio	0%
Valor_em_novo	20%	alt_bonus	0%
tipo_cliente_nvida	17%	idd_veiculo	0%
tipo_cliente_vida	13%	ant_agente	0%
NUM_CC2	1%	tipo_cliente_auto	0%
idade_carta	0%	tipo_cliente_produto	0%
MarcaEmblem	0%	agrava_5anu	0%
NUM_CV2	0%	agrava_1anu	0%
Ano_Subscricao	0%	cartao	0%
Distrito_apol	0%	antig_apol	0%
CANAL_AGENTE	0%	fPagamento	0%
Month_Renewal	0%	Int_Var_Rel_Depois	0%
CVE_TARIFA	0%	Int_Var_Rel_Depois_n1	0%
formacob	0%	CDSUBRAMO	0%

*Tabela 3.14 - 4ª Iteração do teste de Wald*

#### **5º Iteração:**

A variável “tipo\_cliente\_nvida” é a variável com o maior *p-value*, superior a 5%, desta forma esta é a quinta a ser retirada do modelo.

Variável	Wald p Value (Chi-Sq)	Variável	Wald p Value (Chi-Sq)
PESO_BRUTO	6%	alt_premio	0%
CVE_COMBUSTIVEL	4%	alt_bonus	0%
idade_tomador	3%	idd_veiculo	0%
tipo_cliente_nvida	17%	ant_agente	0%
tipo_cliente_vida	13%	tipo_cliente_auto	0%
NUM_CC2	0%	tipo_cliente_produto	0%
idade_carta	0%	agrava_5anu	0%
MarcaEmblem	0%	agrava_1anu	0%
Ano_Subscricao	0%	cartao	0%
Distrito_apol	0%	antig_apol	0%
CANAL_AGENTE	0%	fPagamento	0%
Month_Renewal	0%	Int_Var_Rel_Depois	0%
CVE_TARIFA	0%	Int_Var_Rel_Depois_n1	0%
formacob	0%	NUM_CV2	0%
LABEL_PACK	0%	CDSUBRAMO	0%
categ	0%		

*Tabela 3.15 - 5ª Iteração do teste de Wald*

### **6º Iteração:**

A variável “tipo\_cliente\_vida” é a variável com o maior *p-value*, superior a 5%, desta forma esta é a sexta a ser retirada do modelo.

Variável	Wald p Value (Chi-Sq)	Variável	Wald p Value (Chi-Sq)
PESO_BRUTO	6%	alt_premio	0%
CVE_COMBUSTIVEL	4%	alt_bonus	0%
idade_tomador	3%	idd_veiculo	0%
tipo_cliente_vida	13%	ant_agente	0%
NUM_CC2	0%	tipo_cliente_auto	0%
idade_carta	0%	tipo_cliente_produto	0%
MarcaEmblem	0%	agrava_5anu	0%
Ano_Subscricao	0%	agrava_1anu	0%
Distrito_apol	0%	cartao	0%
CANAL_AGENTE	0%	antig_apol	0%
Month_Renewal	0%	fPagamento	0%
CVE_TARIFA	0%	Int_Var_Rel_Depois	0%
formacob	0%	Int_Var_Rel_Depois_n1	0%
LABEL_PACK	0%	NUM_CV2	0%
categ	0%	CDSUBRAMO	0%

*Tabela 3.16 - 6ª Iteração do teste de Wald*

### **7º Iteração:**

A variável “PESO\_BRUTO” é a variável com o maior *p-value*, superior a 5%, desta forma esta é a sétima a ser retirada do modelo.

Variável	Wald p Value (Chi-Sq)	Variável	Wald p Value (Chi-Sq)
PESO_BRUTO	6%	alt_bonus	0%
CVE_COMBUSTIVEL	4%	idd_veiculo	0%
idade_tomador	3%	ant_agente	0%
NUM_CC2	0%	tipo_cliente_auto	0%
idade_carta	0%	tipo_cliente_produto	0%
MarcaEmblem	0%	agrava_5anu	0%
Ano_Subscricao	0%	agrava_1anu	0%
Distrito_apol	0%	cartao	0%
CANAL_AGENTE	0%	antig_apol	0%
Month_Renewal	0%	fPagamento	0%
CVE_TARIFA	0%	Int_Var_Rel_Depois	0%
formacob	0%	Int_Var_Rel_Depois_n1	0%
LABEL_PACK	0%	NUM_CV2	0%
categ	0%	CDSUBRAMO	0%
alt_premio	0%		

*Tabela 3.17 - 7ª Iteração do teste de Wald*

### 8º Iteração:

Todas as variáveis do modelo têm um valor de *p-value* inferior a 5%. Desta forma, não retiramos mais nenhuma variável ao modelo, e assim o teste de *Wald* termina com 28 variáveis. Sendo o AIC=197280.9.

Variável	Wald p Value (Chi-Sq)	Variável	Wald p Value (Chi-Sq)
CVE_COMBUSTIVEL	5%	alt_bonus	0%
idade_tomador	3%	idd_veiculo	0%
NUM_CC2	0%	ant_agente	0%
idade_carta	0%	tipo_cliente_auto	0%
Ano_Subscricao	0%	tipo_cliente_produto	0%
Distrito_apol	0%	agrava_5anu	0%
CANAL_AGENTE	0%	agrava_1anu	0%
Month_Renewal	0%	cartao	0%
CVE_TARIFA	0%	antig_apol	0%
formacob	0%	fPagamento	0%
LABEL_PACK	0%	Int_Var_Rel_Depois	0%
categ	0%	Int_Var_Rel_Depois_n1	0%
MarcaEmblem	0%	NUM_CV2	0%
alt_premio	0%	CDSUBRAMO	0%

Tabela 3. 18 - 8ª Iteração do teste de Wald

Após aplicar os dois métodos conclui-se que o teste de *Wald* é o que determina o modelo final, pois é este método que apresenta um AIC menor. Desta forma, as variáveis selecionadas pelo teste de *Wald* são as que melhor explicam o modelo de anulação automóvel.

### 3.3.3. Modelo final

Após a construção do modelo de regressão logística com o auxílio do teste de *Wald*, obtemos os valores de  $\beta^{[10]}$  e o desvio padrão:

✓ Desvio padrão = 0.03136

Sendo PA a probabilidade de anulação, temos que o modelo criado é definido pela seguinte expressão:

$$\text{Logit(PA)} = \beta_0 + \beta_1 \text{CVE\_COMBUSTIVEL} + \beta_2 \text{idade\_tomador} + \beta_3 \text{NUM\_CC2} + \beta_4 \text{idade\_carta} + \beta_5 \text{Ano\_Subscrição} + \beta_6 \text{Distrito\_apol} + \beta_7 \text{CANAL\_COMERCIAL} + \beta_8 \text{Month\_Renewal} + \beta_9 \text{CVE\_TARIFA} + \beta_{10} \text{formacob} + \beta_{11} \text{LABEL\_PACK} + \beta_{12} \text{categ} + \beta_{13} \text{MarcaEmblem} + \beta_{14} \text{alt\_premio} + \beta_{15} \text{alt\_bonus} + \beta_{16} \text{idd\_veiculo} + \beta_{17} \text{ant\_agente} + \beta_{18} \text{tipo\_cliente\_auto} + \beta_{19} \text{tipo\_cliente\_produto} + \beta_{20} \text{agrava\_5anu} + \beta_{21} \text{agrava\_1anu} + \beta_{22} \text{cartao} + \beta_{23} \text{antig\_apol} + \beta_{24} \text{fPagamento} +$$

<sup>10</sup> Os valores de  $\beta$  são confidenciais



$$\beta_{25} * \text{Int\_Var\_Rel\_Depois} + \beta_{26} * \text{Int\_Var\_Rel\_Depois\_n1} + \beta_{27} * \text{NUM\_CV2} + \beta_{28} * \text{CDSUBRAMO}$$

### 3.4. Avaliação do modelo

Neste subcapítulo, pretendemos verificar o quão bem ajustado o modelo se encontra. Para isso, no momento de criação da base de dados, esta foi dividida aleatoriamente em base modelo (70%), base utilizada para modelar, e em base de teste (30%). É a partir desta base teste que vamos testar o modelo.

#### 3.4.1. Curva de ROC

De forma a avaliar a capacidade discriminatória do modelo calculámos a área abaixo da curva de ROC, designada por AUC (Area Under The Curve). Quanto maior for o valor do AUC melhor o modelo explica os dados.

Nas imagens abaixo podemos ver as curvas de ROC para a base do modelo e para a base de testes. O AUC da base modelo têm um valor de 0.763 enquanto o AUC da base teste é um pouco inferior, tendo um valor de 0.747, estes valores indicam-nos que o modelo tem uma capacidade discriminatória aceitável. Não dando assim nenhum “sinal” de que o modelo esteja desajustado.

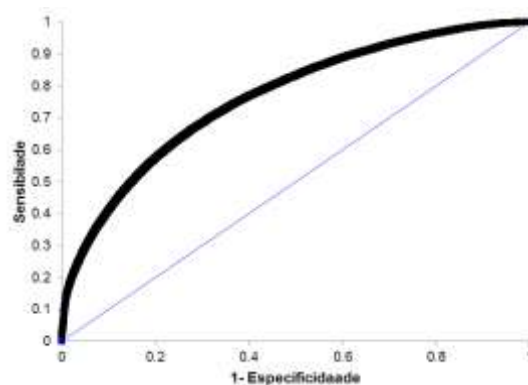


Gráfico 3.41 – Curva de ROC da base modelo

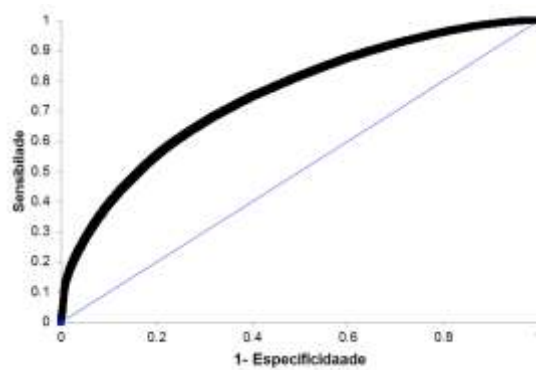


Gráfico 3.42 – Curva de ROC da base teste

### 3.4.2. Sensibilidade e Especificidade

A partir dos diferentes pontos de corte é possível construir os gráficos apresentados abaixo. Através dos gráficos verificamos que o ponto de corte perfeito das duas bases (base do modelo e base de testes) é 0.15.

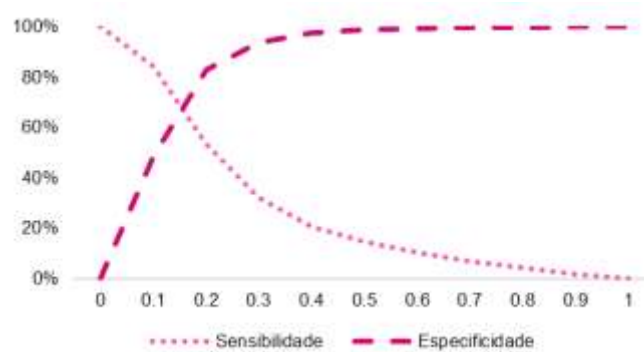


Gráfico 3.43 – Sensibilidade e Especificidade da base modelo

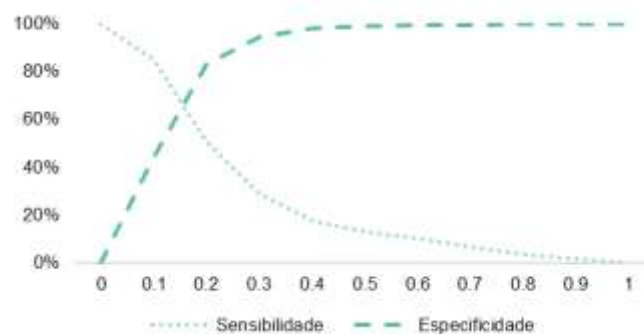


Gráfico 3.44 - Sensibilidade e Especificidade da base teste

Através dos vários valores do ponto de corte é possível construir a matriz de confusão. Nas tabelas abaixo podemos verificar as matrizes de confusão da base do modelo e da base teste respetivamente.

		Observados		Total
		Positivos	Negativos	
Previstos	Positivos	31493	77028	<b>108521</b>
	Negativos	14893	184169	<b>199062</b>
Total		<b>46386</b>	<b>261197</b>	<b>307583</b>

*Tabela 3.19 – Matriz confusão da base modelo*

		Observados		Total
		Positivos	Negativos	
Previstos	Positivos	13276	34542	<b>47818</b>
	Negativos	6690	77142	<b>8832</b>
Total		<b>19966</b>	<b>111684</b>	<b>131650</b>

*Tabela 3.20 – Matriz de confusão da base teste*

Através das matrizes de confusão é possível calcular a precisão do modelo, sendo a  $Precisão = \frac{VP+VN}{P+N}$ . Então na nossa base modelo a precisão é de 70.1% e da base de testes é de 68.7%.

Da mesma forma que verificamos nos valores do AUC, também os valores de precisão da base do modelo e de testes são muito idênticos, no entanto, a percentagem de apólices bem classificadas é mais alta na base do modelo. Com estes valores de precisão é possível afirmar que as apólices do modelo estão bem classificadas.

## CAPÍTULO IV – Conclusão

A realização deste projeto traduz-se numa aplicação dos conhecimentos adquiridos durante o mestrado à realidade profissional, neste caso no mercado segurador. Sendo a AGEAS uma empresa internacional, com anos de experiência, foi uma honra poder desenvolver este projeto na companhia. Espero que traga todos os benefícios previstos e que proporcione o aumento da rentabilidade assim como o aumento da satisfação do cliente.

Neste projeto utilizou-se a regressão logística para modelar a probabilidade de anulação automóvel (no processo de renovação), com a função variação do prémio e outras variáveis. Avaliou-se a capacidade de o modelo prever corretamente o perfil das apólices propícias a anularem na renovação.

Este modelo permite calcular a probabilidade de anulação estimada para cada cliente por intervalo de variação de prémio. Desta forma, é possível obter o perfil de clientes que apresentam maior elasticidade à variação de prémio e no contexto do negócio é possível implementar decisões mais acertadas no processo de Repricing, contribuindo para a rentabilidade do negócio.

A fase mais longa do projeto foi a construção da base de dados. Esta é fundamental para modelizar, por isso, é essencial que esteja correta e bem desenvolvida.

As variáveis incluídas no modelo fornecem informação sobre as diferentes reações dos clientes durante o período de renovação, no entanto, os resultados obtidos vão ao encontro das nossas expectativas. Do ponto de vista do negócio, algumas variáveis do modelo são mais essenciais do que outras - por exemplo, variação de prémio, forma de pagamento, forma de cobrança, são essenciais no processo de *repricing*.

No gráfico abaixo podemos ver a reação do cliente face à variação do prémio na renovação anterior e na renovação atual. Tal como é expectável, os clientes que mais anulam são aqueles que sofrem dois aumentos em duas renovações consecutivas. É fundamental perceber que tipo de clientes temos neste grupo, uma vez que não é intenção da companhia perder os “bons” clientes.

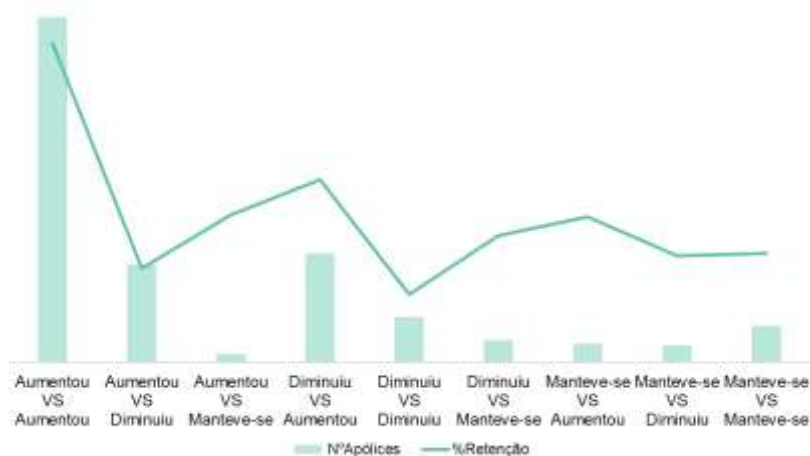


Gráfico 3.5 – Alteração de prémio nas duas anuidades

Concluimos também que o modelo apresenta uma capacidade de previsão global razoável, medida pelo AUC e pelo valor de precisão.

Após alcançarmos os objetivos pretendidos, podemos ainda dar recomendações para o futuro. Sabemos que o nível de satisfação do cliente em relação à companhia é muito importante no momento de renovação da apólice, desta forma, seria interessante incluir no modelo uma variável do nível de satisfação do cliente ou mesmo informações sobre possíveis reclamações feitas pelo cliente no passado. Por outro lado, as inevitáveis mudanças no ambiente económico e de negócios, bem como o constante crescimento da carteira de automóvel, implicam que a qualidade do modelo possivelmente diminuirá ao longo do tempo. Desta forma, é fundamental recalibrar o modelo anualmente.

Apesar deste projeto abordar o mesmo tema do projeto da Joana Garraio, Modelação da Taxa de Anulação no Seguro Automóvel (2015), os projetos diferem em alguns aspetos. Ambos foram feitos na mesma companhia, no entanto, devido à elevada competitividade dos mercados, os processos de *repricing* sofreram bastantes alterações nos últimos anos, sendo que atualmente existe um foco maior no cliente. Este modelo foi construído através do software Emblem, usando o método GLM, regressão logística, enquanto no projeto Modelação da Taxa de Anulação no Seguro Automóvel (2015), o software utilizado foi o SAS. O presente modelo contempla mais variáveis, algumas delas externas à companhia. O universo em análise dos dois modelos também difere, no projeto “Modelação da Taxa de Anulação no Seguro Automóvel” apenas se retirou as apólices frotas.

Estes modelos são fundamentais para as seguradoras e embora a maioria dos trabalhos se concentrem em clientes automóveis individuais, a mesma abordagem pode ser aplicada a outras linhas de negócio, como o seguro de multirriscos habitação.

Futuramente este modelo vai ser aplicado no software Radar, de forma a apoiar e desenvolver estratégias de *repricing* nas renovações

## Referências bibliográficas

- [1] Ageas Portugal (outubro 2016). Apresentação da empresa. [https -  
//www.Ageas.pt/documents/11601/76264134/Apresentacao\\_Institucional\\_Ageas.pdf/94  
e14912-7253-4a01-a26f-cef27fa19cb5](https://www.Ageas.pt/documents/11601/76264134/Apresentacao_Institucional_Ageas.pdf/94e14912-7253-4a01-a26f-cef27fa19cb5)
- [2] História dos seguros em Portugal (outubro 2016). [http -//historiadoseguro.com/historia-  
de-seguros-portugal/](http://historiadoseguro.com/historia-de-seguros-portugal/)
- [3] Autoridade de Supervisão de Seguros e Fundos de Pensões (outubro 2016). O que é o contrato de seguro. [http -//www.asf.com.pt/NR/exeres/E527B387-8D5D-4701-8143-  
EF1EF088CC56.htm](http://www.asf.com.pt/NR/exeres/E527B387-8D5D-4701-8143-EF1EF088CC56.htm)
- [4] Alpuim, T, Modelos Lineares - Notas de apoio à disciplina, 2015.
- [5] Esteves, R, Atividade Seguradora - Notas de apoio à disciplina, 2016.
- [6] Figueiredo, P, Atividade Seguradora - Notas de apoio à disciplina, 2016
- [7] Garraio, J. (2015). Modelação da Taxa de Anulação no Seguro Automóvel. Tese de mestrado, Universidade de Lisboa – Faculdade de Ciências.
- [8] Mousinho, G. (2016). Modelling Renewal Price Elasticity - An Application To The Motor Portfolio Of Ocidental. Tese de mestrado, Universidade de Lisboa – Instituto Superior de Economia e Gestão.
- [9] Faustino, B. (2016). Modelação e Previsão no Seguro Automóvel. Tese de mestrado, Universidade do Porto – Faculdade de Ciências.
- [10] Ginja, A. (2017). Modelação da Zona Tarifária no Seguro Automóvel. Tese de mestrado, Universidade de Lisboa – Faculdade de Ciências.
- [11] Daniel, L. (2009). ABC dos Seguros. Vida Economica -Editorial SA.
- [12] Gilberto, F. (2008). Manual Prático dos Seguros. Lidel.
- [13] Capital mínimo obrigatório RC (janeiro 2018) - [http://netseguros.pt/aumento-capital-  
minimo-obrigatorio/](http://netseguros.pt/aumento-capital-minimo-obrigatorio/)
- [14] De Jong, P. and Heller, G. Z. (2009). Generalized Linear Models for Insurance Data, 1st ed. Cambridge: Cambridge University Press.
- [15] Associação Portuguesa de Seguradores (setembro 2019). Panorama do Mercado Segurador 14/15. –  
[https://www.apseguradores.pt/Portal/Content\\_Show.aspx?ContentId=2248&PageId=8&  
MicrositeId=1&CategoryId=70](https://www.apseguradores.pt/Portal/Content_Show.aspx?ContentId=2248&PageId=8&MicrositeId=1&CategoryId=70)
- [16] Associação Portuguesa de Seguradores (setembro 2019). –  
<http://www.asf.com.pt/NR/exeres/BF408252-BF68-4555-8AD1-34E31950AE11.htm>